PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

JAN 14 1936

VOL. XXVII NOVEMBER, 1936 NO. 8

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

CONTENTS

A Quantitative Study of the Major Interests of College Students . . 561
SISTER ANNETTE WALTERS AND ALVIN C. EURICH

The Associational Aspect of Musical Responses in School Children 572
J. H. HIGGINSON

The Problem of Principal Components: II—The Argument for Communalities 581
C. E. KELLOGG

An Experiment in Comparing Responses to Two Types of Poetry, Rhymed and Unrhymed, in Nursery School, Kindergarten, and First Grade 591
DOROTHY BARUCH

The Definition of Learning 603
JOHN NOBLE WASHBURNE

Quantitative Wholes: Classroom Experiments 612
MAY V. SEAGOE

The Speed Factor in Reading Performance 621
VERNA L. ANDERSON AND MILES A. TINKER

The Residual Index 625
C. W. YOUNG

Racial Comparisons in Performance on the American Council Psychological Examination 630
T. M. LIVESAY

Book Reviews 635

Publications Received 637

$6.00 per Year · Published Monthly September to May

WARWICK & YORK, INC.

BALTIMORE, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Established 1910

EDITORS

JACK W. DUNLAP
Fordham University

HAROLD E. JONES
University of California

PERCIVAL M. SYMONDS
Teachers College,
Columbia University

H. E. BUCHHOLZ, *Managing Editor*

THE price of the Journal is $6.00 a year in the United States; $6.40 in foreign countries. Part-year subscriptions are 90 cents for each number ordered. Back volumes are $7.00 each; back issues are $1.10 each except when more than five years old, and then $1.20 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Unsolicited manuscripts should be accompanied with return postage. Manuscripts, books and other materials for review, and correspondence regarding editorial and business matters should be addressed to the Publishers.

WARWICK AND YORK · *Publishers* · BALTIMORE, MD.

THE ACTIVITY MOVEMENT

By CLYDE HISSONG

In an attempt to overcome the weaknesses of the traditional school organization many progressive schools have developed new programs. These programs are so similar in character that collectively the changes have been referred to as the activity movement. This movement has claimed the center of the educational stage for a length of time sufficient to have engendered widespread interest in its outcomes and in its basic philosophy.

In Doctor Hissong's study an attempt has been made to discover the principles underlying the present activity movement, to determine the influence of traditional concepts in shaping the trends of the movement, and to see if in the light of the present knowledge of the child and his relation to his environment the movement rests upon a justifiable basis.

$2.00 plus 10¢ postage

WARWICK AND YORK

Publishers BALTIMORE

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXVII November, 1936 Number 8

A QUANTITATIVE STUDY OF THE MAJOR INTERESTS OF COLLEGE STUDENTS

SISTER ANNETTE WALTERS

The College of St. Catherine

AND

ALVIN C. EURICH

University of Minnesota

The purpose of this study was to develop interest scales which may be used in predicting the major interests of college students before they have chosen their fields of specialization. The value of such scales for predicting scholastic success has been suggested by the studies of Wyman[1] and Langlie,[2] and the possibility of developing them has been demonstrated experimentally by Van Tuyl and Eurich.[3] It is assumed that differences in interest may be responsible for the discrepancy between aptitude test scores and college grades, and that an interested student will accomplish far more than a less interested student of equal ability. If to the already existing instruments for guiding incoming freshmen into the proper curricula, an interest inventory of demonstrated importance in selecting subject-matter fields can be added, academic mortality due to lack of motivation may be materially decreased.

---

[1] Wyman, J. B.: "Tests for intellectual, social, and activity interests." In L. M. Terman *et al.*, *Genetic Studies of Genius.* Stanford University Press, Vol. I, 1925, pp. 445–483.

[2] Langlie, T. A.: "Interests and scholastic proficiency." *Personnel Journal,* Vol. IX, 1930–1931, pp. 246–250.

[3] Van Tuyl, K. and Eurich, A. C.: "Measuring the interests of college students with different major subjects." *Journal of Applied Psychology,* Vol. XVIII, 1934, pp. 27–44.

CHOICE OF SUBJECTS AND PROCEDURE

The Minnesota Interest Blank, consisting of one hundred occupational items and one hundred general items such as personal characteristics, amusements, educational subjects, books, and magazines, was given to a total of four hundred twenty-six women students at The College of St. Catherine, winter quarter, 1935. Test blanks of upper classmen majoring in chemistry, biology, dietetics, and mathematics were then sorted out and classified as the science-mathematics group. Test blanks of upper classmen majoring in history, social science, and child welfare were combined to form the social science division. English majors were not combined with any other groups. The criterion groups thus obtained were as follows:

Major Subject	*N*
Science-mathematics	64
Social science	67
English	59

Responses made to each of the two hundred items listed on the interest inventory were tabulated according to the criterion groups represented, and the percentage of each of the three groups crossing out the symbols *L*, *D*, *O*, or *U* after an item was determined. Crossing out the letter *L* indicated a liking for that item. Crossing out *D*, *O*, or *U* indicated, respectively, dislike, indifference, or ignorance. For each item the significance of the difference between the percentages of any two major groups responding in the same way was determined by calculating the ratio of the difference to its probable error. Any difference which was equal to or greater than three times the probable error of the difference was considered significant. Theoretically, items satisfying this criterion would be expected to yield differential responses in ninety-six out of every one hundred comparisons. Items yielding significant differences between one major group and one or both of the others were used to build up a scoring key for that group. Items showing significant differences between one major group and two others were given a weight of two. The total interest inventories of the total four hundred twenty-six students were then scored and analyzed:

1. To determine the extent to which they would differentiate the criterion groups on which they had been derived;

2. To determine the extent to which they would differentiate groups of freshmen which had evidenced preferences for certain major fields but which had not definitely chosen major subjects;

3. To determine the relation of major interest scores to achievement in the associated subjects as measured by a battery of comprehensive tests;[1]

4. To determine the significance of differences between responses of freshmen and seniors.

An example of the procedure by which scoring keys were derived is given below:

Item and symbol	Group I science-mathematics		Group II social science		Group III English	
	Per cent crossing out	PE²	Per cent crossing out	PE²	Per cent crossing out	PE²
Linguist.						
L.....	39	.001689	37	.001576	64	.001747
D.....	22	.001225	30	.001421	14	.000918
O.....	30	.001475	25	.001274	19	.001176
U.....	9	.000586	7	.000437	3	.000219

Groups I and II			Groups I and III			Groups II and III		
$\sqrt{\Sigma PE_s^2}$	*D*	D/PE_D	$\sqrt{\Sigma PE_s^2}$	*D*	D/PE_D	$\sqrt{\Sigma PE_s^2}$	*D*	D/PE_D
	2	...	.0586	25	4.3*	.0651	27	4.1*
.0514	8	1.6	.0463	8	1.7	.0484	16	3.3*
.0524	5	1.0	.0515	11	2.1	.0495	6	1.2
	2	...	.0284	6	2.1	.0256	4	

* Statistically significant difference.

It will be observed that the percentage of English majors responding "*L*" to the item "linguist" is significantly g ater than the percentage of science-mathematics $\left(\text{III} - \text{I} - \frac{D}{PE_D} = 4.3\right)$ or social

[1] This battery consisted of a comprehensive general culture test, comprised of three sections, namely, history and social science, foreign literature, and fine arts, and of a comprehensive general science test. These tests were used in the annual sophomore testing program of the American Council on Education.

science $\left(\text{III} - \text{II} - \frac{D}{\text{PE}_D} = 4.1\right)$ majors responding in that way. Since this symbol differentiated the English group from the other two major groups, it was given a weight of two in the English scoring key. The response "*D*" to the same item was made a significantly greater number of times by the social science majors than by the English majors ($D/\text{PE}_D = 3.3$). Since, however, it did not differentiate the science-mathematics from the social science group, it was given a weight of only one in the social science scoring key.

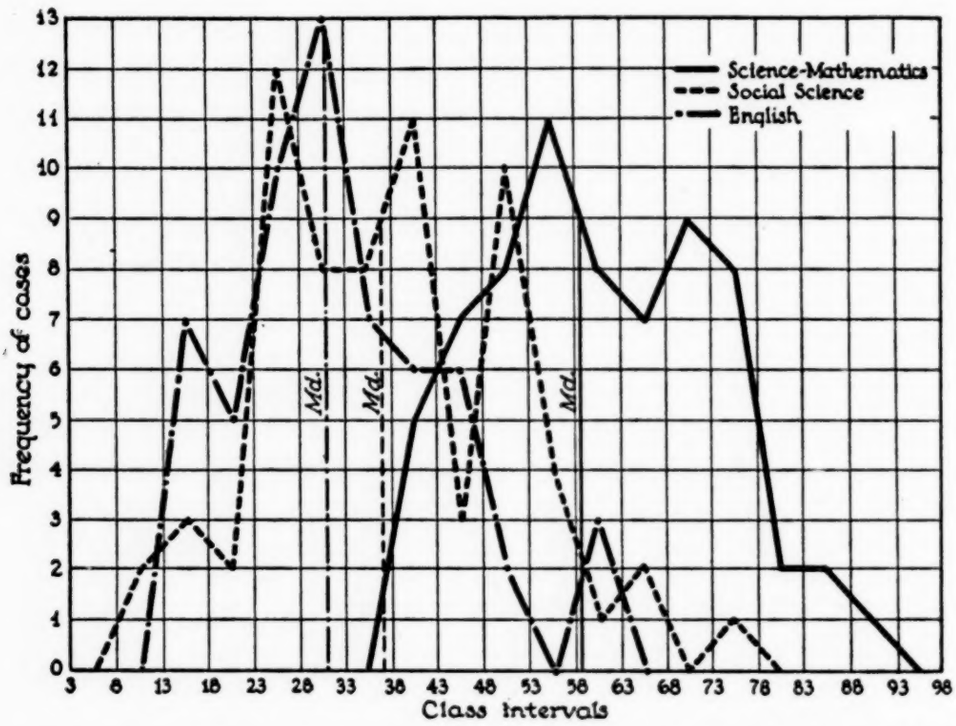

FIG. 1.—Scores made by science-mathematics majors, social science majors, and English majors on the science-mathematics scale.

RESULTS AND DISCUSSION

By the technique described, eighty-one items were obtained which differentiated the science-mathematics from the social science group, ninety-eight items which differentiated the science-mathematics from the English group, and sixty-one items which differentiated the social science from the English group. Interest scales for each of these three major groups were constructed from the items which differentiated a given major group from one or both of the others, and each item in the scales was weighted according to the number of major groups of which it was differential. These interest scales successfully

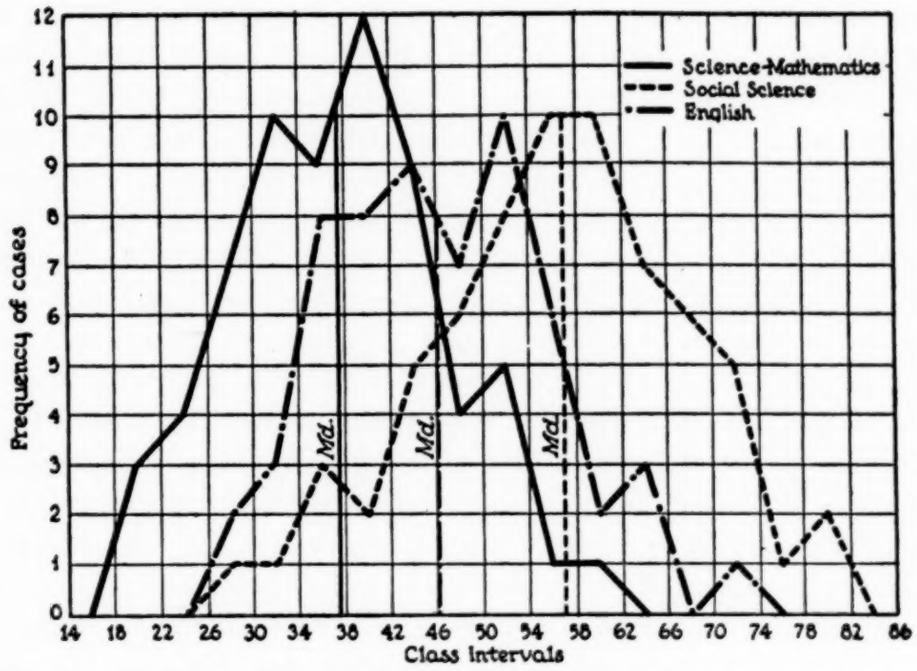
Science-Mathematics
Social Science
English
Md.
Md.
Md.
Frequency of cases
Class intervals
0 1 2 3 4 5 6 7 8 9 10 11 12
14 18 22 26 30 34 38 42 46 50 54 58 62 66 70 74 78 82 86

FIG. 2.—Scores made by social science majors, English majors, and science-mathematics majors on the social science scale.

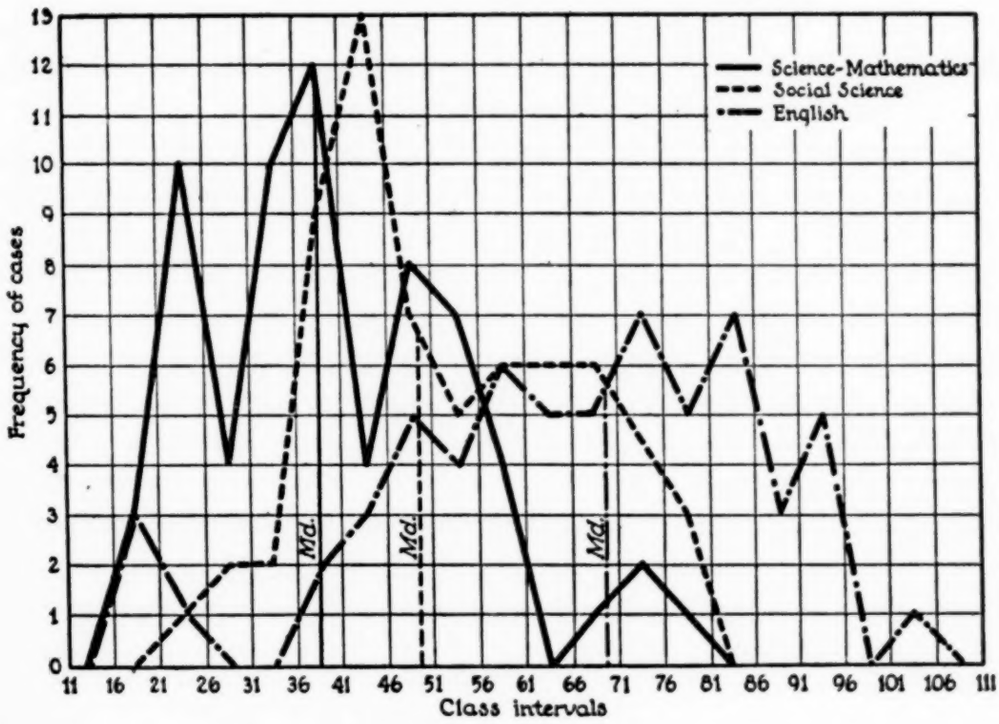
Science-Mathematics
Social Science
English
Md.
Md.
Md.
Frequency of cases
Class intervals
0 1 2 3 4 5 6 7 8 9 10 11 12 13
11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 111

FIG. 3.—Scores made by the English majors, social science majors, and the science-mathematics majors on the English scale.

differentiated the various criterion groups, as evidenced by the significant differences between means and the small percentage of overlapping between groups. On the science-mathematics scale, the mean scores obtained were as follows: Science-mathematics group, 56.48; social science group, 31.91; and English group, 30.87. Mean scores for the three groups on the social science scale were: Social science, 53.11; English 44.57; and science-mathematics, 33.35. Mean scores on the English scale were: English, 64.48; social science, 45.03; and science-mathematics, 34.84. The difference between the means divided by the standard error of the difference was for the science-mathematics and social science groups on their respective scales, 11.32 and 11.07; for the science-mathematics and English groups on their respective scales, 12.55 and 11.36; and for the social science and English groups on their respective scales, 4.62 and 6.63. All of these data indicate statistically reliable differences. The amount of overlapping in terms of the median between the science-mathematics and social science groups on their respective scales, was six and three per cent, between the science-mathematics and English groups, five and five per cent, and between the social science and English groups, ten and fifteen per cent. These results are summarized in Figs. 1, 2, and 3.

Odd-even reliability coefficients, corrected by the Spearman-Brown formula are given in Table I. These coefficients are as high as those usually reported for very reliable tests.

Freshmen groups which had evidenced preferences for certain major fields but which had not definitely decided upon major subjects were differentiated by significant mean differences in four out of six comparisons, as shown in Table II. Significant differences were obtained between the science-mathematics and the social science groups on the social science scale, between the science-mathematics and English groups on their respective scales, and between the social science and English groups on the English scale. The percentage of overlapping between these groups is greater than that between criterion groups, except that there is no overlapping between the science-mathematics

TABLE I.—ODD-EVEN RELIABILITY COEFFICIENTS

Scale	*N*	*r*	r_x
Science-mathematics	184	.809	.894
Social science	184	.876	.934
English	184	.719	.837

TABLE II.—RELIABILITY OF DIFFERENCES BETWEEN OBTAINED MEANS OF FRESHMAN WHO HAVE TENTATIVELY CHOSEN SCIENCE-MATHEMATICS, SOCIAL SCIENCE, OR ENGLISH MAJORS

Scale	Major group									Reliability of differences					
	Group I science-mathematics (N = 40)			Group II social science (N = 32)			Group III English (N = 32)			Groups I and II			Groups I and III			Groups II and III	
	Mean	SD	SE_m	Mean	SD	SE_m	Mean	SD	SE_m	$SE_{D_{m_1-m_2}}$	D/SE_D		$SE_{D_{m_1-m_2}}$	D/SE_D		$SE_{D_{m_1-m_2}}$	D/SE_D
Science-mathematics	45.9	13.13	2.08	40.7	11.23	1.95	30.3	19.68	1.71	2.87	1.81		2.69	5.79*			
Social science	40.5	10.43	1.65	48.6	10.47	1.85	48.7	18.83	1.56	2.48	3.26*		....	....		2.43	.04
English	44.7	19.00	3.00	49.3	12.52	2.21	68.1	14.42	2.55	....	....		3.94	5.94*		3.37	5.58*

* Statistically significant difference.

and English groups on the science-mathematics scales, as shown in Table III.

Of the three interest scales, the science-mathematics scale appears to be most valid, and the social science scale least valid, as determined by the method of overlapping. These results are in harmony with those obtained by Van Tuyl and Eurich.[1] There is greater similarity between interests of social science majors, as represented in this study, and English majors, than between any other criterion groups.

Interests scores of groups of upper classmen majoring in subjects other than science-mathematics, social science, or English, are given in Table IV. From this Table it is apparent that students majoring in foreign languages are similar in their interests to English majors, since they score consistently high on the English scale and consistently low on the science-mathematics scale. A similar trend is observed for music majors. The latter group, however, is so small that one cannot have undue confidence in any conclusions which might be drawn from

TABLE III.—AMOUNT OF OVERLAPPING BETWEEN DIFFERENT FRESHMEN GROUPS

Scale	Group I science-mathematics median	Group II social science median	Group III English median	Percentage of overlapping between groups I and II	Percentage of overlapping between groups I and III	Percentage of overlapping between groups II and III
Science-mathematics.....	49.66	38.83	30.50	18	0	
Social science.....	42.00	51.42	47.33	24	..	40
English.....	47.00	53.00	62.16	..	20	12

TABLE IV.—INTEREST SCORES OF UPPER CLASSMEN MAJORING IN SUBJECTS OTHER THAN SCIENCE-MATHEMATICS, SOCIAL SCIENCE, OR ENGLISH

Major subject	*N*	Mean score on science-mathematics scale	Mean score on social science scale	Mean score on English scale
Foreign languages.....	21	34.42	43.52	62.90
Music.....	13	43.61	48.84	62.15
Art.....	8	38.75	51.00	50.6
Physical education.....	14	45.43	48.00	46.71

[1] *Op. cit.*

its scores. The results for other major groups are also inconclusive because of the small number of cases.

TABLE V.—CORRELATION BETWEEN MAJOR INTERESTS AND PERCENTILE RATINGS ON THE MINNESOTA COLLEGE ABILITY TEST

Scale	*N*	*r*	PE
Science-mathematics	390	−.06	±.03
Social science	390	−.10	±.03
English	390	+.19	±.03

Relative interests as measured by the three scales developed in this study appear to be independent of intelligence ratings, since the distribution of scores for the different major groups remained the same after the groups had been equated for intelligence. Moreover, the science-mathematics and the social science scales yielded slightly negative correlations with intelligence as measured by the Minnesota College Ability Test. The English scale yielded a low positive correlation with intelligence. These figures are summarized in Table V.

English interests were found to yield a slight positive correlation with scores on the fine arts, foreign literature, and literary acquaintance tests. In view of the fact that this scale correlates equally well with intelligence, however, these results cannot be attributed to interest alone. There is no evidence that this scale has any capacity for predicting scholarship in the field of English. A zero correlation is obtained between social science interests and scores on the history and social science tests. All that can be expected, therefore, is that major interest scores obtained on these two scales, when considered along with other factors, will be of value in placing the freshman in a curriculum which will best satisfy his dominant interests and drives. The science-mathematics scale, which was the most valid in differentiating between groups, is positively correlated with achievement in general science, as measured by the Coöperative General Science Test. The correlation, though low, is statistically significant, and since this scale bears a zero relationship to intelligence, the correlation may be attributed to interest alone. Hence, it may be concluded that the latter scale is capable of differential prediction not only of interests but also of achievement. These results are given in Table VI.

When responses to each of the two hundred items on the interest inventory were tabulated separately for freshmen and seniors, and the

TABLE VI.—RELATION OF MAJOR INTEREST TO ACHIEVEMENT

Achievement test	Interest scale	*N*	*r*	PE
Fine arts	English	251	.193	± .041
Foreign literature	English	251	.184	± .041
Literary acquaintance	English	251	.164	± .041
History and social science	Social science	389	.047	± .034
General science	Science-mathematics	253	.297	± .038

results converted into percentages, forty-three items were found to yield statistically significant differences, twenty-one of which represented occupational interests. The results indicate a greater liking on the part of seniors for occupations, particularly occupations on a relatively high social economic level, than is evidenced by freshmen, although the occupational background of the two groups is approximately the same. They also indicate a greater preference or a less marked dislike for certain school subjects. Greater interest in social science periodicals is evidenced by seniors. Symphony concerts, English literature, and a number of specific books are more liked by seniors than by freshmen.

To determine the extent to which the results described above were conditioned by differences in intelligence, each senior was matched with a freshman having the same percentile rating on the Minnesota College Ability Test, or differing from it by one or two percentile points. An analysis of the interest inventories of these two groups yielded only twenty-three differential responses. The trend of the differences is similar to that of the non-equated groups, although the number of differences is reduced nearly one-half. If our hypothesis is correct, namely, that the present seniors were, four years ago, what the freshmen of equal intelligence are now, the results argue for a remarkable permanence of interests. On the other hand, they suggest that important changes are effected as the result of increasing maturity and the pursuance of a liberal arts curriculum. The effect of both of these factors is evidenced by the more favorable attitude toward occupations, and by the greater breadth of educational and vocational interests exhibited by seniors in contrast to freshmen. The fact that seniors in general favor social science periodicals more than freshmen reflects the influence of an education specifically focused upon the development of interest in current problems. Nevertheless, the results strongly suggest that relative interests of college students, as measured by the Minnesota Interest Blank, are not changed to any marked degree by

collegiate experience, and that by studying the interests of the freshman, we may predict to a limited extent what his relative interests as an upper classman will be.

It is possible that the crudeness of the interest inventory has obscured some of the interest changes which might have taken place between the freshmen and senior years. Future research, therefore, must be directed toward the development of an adequate technique for measuring the intensity of an interest as well as its presence or absence.

SUMMARY

(1) Major interest scales were developed for the Minnesota Interest Blank which successfully differentiated the three criterion groups, namely, the science-mathematics, the social science, and the English group, on which they had been derived.

(2) Freshmen groups which had evidenced preferences for certain major fields were likewise differentiated by significant mean differences in four out of six comparisons. In the other two comparisons, the trend was in the same direction as that of the criterion group.

(3) There is a greater similarity of interests between social science majors, as represented in this study, and English majors, than between any other two criterion groups.

(4) Students majoring in foreign languages and in music have interests similar to English majors.

(5) The science-mathematics and social science scales yield a slightly negative correlation with intelligence. The English scale yields a low positive correlation. In both cases the correlation approaches zero.

(6) English interests and social science interests tend to yield a zero correlation with achievement in related fields. Science-mathematics interests are positively correlated with achievement in general science. The correlation, though low, is statistically significant.

(7) Significant differences between responses of freshmen and seniors were obtained on only forty-three items out of the two hundred. When the groups were equated for intelligence, only twenty-three differential responses were obtained. These results suggest that relative interests, as measured by the Minnesota Interest Blank, are remarkably permanent during the college years.

THE ASSOCIATIONAL ASPECT OF MUSICAL RESPONSE IN SCHOOL CHILDREN

J. H. HIGGINSON

Queen Elizabeth Grammar School, Kirkby Lonsdale, Westmorland, England

NATURE OF THE PROBLEM

The path of investigators in the psychology of music is beset with terms—"musical talent," "musicality," "l'intelligence musicale," "musical ability," "musicability," "musical aptitude," and recently, "musical response." Recent researches, edited by Schöen,[1] favor this latter term because of its implication of the unity of cognition, affectivity, and conation in a musical experience. Early investigators, in their efforts to devise a scientific measure, used the isolated fundamentals of musical structure for diagnostic purposes. In one important respect results inferred from a synthesis of reactions to the isolated fundamentals of musical structure are inadequate. They ignore the affective experience integrated in all fully musical experiences.

The newer conception of musical response recognises the integral part of affectivity in any musical experience. Less serious attention was paid formerly to the affective aspect because of the difficulties of approaching it scientifically. "So fascinating an area of research," writes Bingham,[2] "would not have lain virtually unexplored throughout the years, were it not for the baffling complexity which it opposes to any scientific approach." The writer[3] has endeavored to meet these problems and to take the work of the contributors to *The Effects of Music* a step further. Their subjects were nearly all college or university students. The writer's efforts were to glimpse a cross-section of "music in the making." His subjects were boys ranging from ten to fourteen years of age. They were children from average homes—in embryo "the average man," "the man in the street" who forms the bulk of our population.

---

[1] Schöen, Max: *The Effects of Music.* Series of psychological researches by various investigators. Publ. International Library of Psychology, Philosophy and Scientific Method.

[2] Bingham: *The Effects of Music.* p. 2.

[3] *An Experimental Investigation of the Musical Response of School Children.* Master of Education thesis, accepted by the University of Leeds, England.

STATEMENT OF THE PROBLEM

To map the whole affective experience is beyond the scope of a single paper. One aspect, the associative, is here offered. Associations in a musical experience with children represent one aspect of the affective state evoked by the composition. They allow a three-fold interpretation. One type of association clearly reveals origin in an affective state corresponding to the inherent qualities of the composition. A second type indicates perceptual activity. Thus "it sounded like a horse *trotting*" (Merrymaker's Dance), "it made me think of Vienna and gondolas *gliding* on the water" (Spring Song), "a long *line* of men *moving* very *slow*" (Asa's Death). Such associations indicate a perception of the predominant characteristic of the music. These associations represent a translation of the perception into word-clothing. The third type of association differs from the foregoing and is musically irrelevant. This type of association is produced when attention and interest in the stimulus have dwindled. Wandering trains of thought are born and developed for their own sake irrespective of the musical composition. The first two types are our chief concern in this paper.

C. S. Myers[1] has investigated the connection between musical enjoyment and the tendency to form associations. As he was only dealing with fifteen adult subjects his conclusions may be regarded as tentative. He found an absence of associations in (*a*) the most unmusical, (*b*) frequently in professional musicians, (*c*) in the very highly musical. In (*a*), "the grossly unmusical, music evokes no associations because it evokes no corresponding emotion." Obviously, (*a*) will be applicable to boys of ten to fourteen years; (*b*) not so, and occasionally (*c*). Our problem, then, is to ascertain the nature of the associations formed by boys ten to fourteen years when presented with unknown musical stimuli of various kinds. Are these associations sparsely formed or frequently? Do they indicate (*a*) evoking of an emotional state relevant to the musical stimulus, (*b*) the perceptual activity relevant to the musical stimulus, (*c*) merely wandering trains of thought? What, if any, is the link between enjoyment, the conative factor, and of associativeness? Is the cinema a potent *aural* factor in the emotional and imaginative life of the modern child?

---

[1] Myers: "Individual Differences in Listening to Music." *Brit. Jour. of Psychology*, Vol. VII, 1914, pp. 68–111.

EXPERIMENTAL METHOD

The study of associations was, of course, only part of the work on the much larger mapping of the affective experience. After much preliminary experimentation six gramophone recordings of distinctive types of music were selected:

1. Merrymaker's Dance. Edward German. Nell Gwynn Dances.
2. Quartet in D minor. Schubert. 2nd Movement. First section—so-called "Death and the Maiden" movement was used.
3. Spring Song. Mendelssohn.
4. Storm Passage from William Tell. Rossini.
5. Asa's Death. Grieg.
6. Steel Foundry. From "Symphony of the Machines." Mossolow.

Only instrumental recordings were used, as vocal work introduces a mixed medium which would have added pointless complication to the enquiry. For every record an average of two hundred ten listeners was used—boys ranging from ten to fourteen years of age. After listening to each composition once, the boys were asked to participate in "this new kind of game" by completing a questionnaire which covered six aspects of the affective reaction. One question, "Did it make you think of anything?" was aimed specifically at catching associational effects, but each of the twelve hundred sixty questionnaires was carefully studied for all data that might have interpretative significance in this connection. The underlying principle of this study of associations was kin to that branch of Chinese painting where the Emperor chose a line of poetry, *e.g.*

"One speck of red among ten thousand of green" and left the artists to interpret and depict their own associations. Here the stimulus was verbal, their medium of expression chiefly painting. In the present enquiry the stimulus was music (no verbal titles were ever given), and the medium of expression verbal, and in two experiments, crayon drawing.

DISCUSSION OF ASSOCIATIONS

The order of "associativeness" of the stimuli was:

Record	Per Cent Reporting Associations
Steel Foundry	89
Asa's Death	87
Merrymaker's Dance	72.4
Spring Song	71
"Death and the Maiden"	70
Storm Passage	66.8

To survey the effects of the six stimuli separately:

(*a*) *Steel Foundry.*—This record evoked the highest number of associations. Analysis of these revealed that they were rarely relevant—if we accept the composer's title as a criterion of intended musical communication. As at least half the boys had had first-hand experience of steel foundries, forges and steel hammers, this result was interesting. A typical half dozen associations were: "Dog fight and jumble of cat fight," "Cats in chorus," "A cat going through a ringer," "An old moter," "Barking of dogs and hooters on cars," "A train."

More than any of the other stimuli it produced "wandering trains of thought." The main emotional qualities to be detected in the associations were disgust, anger, amusement. Later, the writer played the "Steel Foundry" to a group of boys fourteen to fifteen year olds who had not been asked to express verbally their affective reaction from the record. As soon as the playing finished, this group was requested to *draw* whatever they thought the record was like. Only one boy drew a steel-hammer (the only subject in the whole testing, verbal and pictorial, to express the composer's title). The remainder connected the record with trains going over long bridges, through tunnels, over echoing arches.

(*b*) *Asa's Death.*—This was the least popular record. It roused a small conative effect. Yet it had a high "associativity." The associations reported from this stimulus were more uniform than for any of the other records. The characteristic words were "death," "funeral," "burial." The most frequent emotional state was of distress. Many instances also occurred of fear, the outcome of the instinct to escape; typical associations were, with "a bogey in a house," "ghosts in a graveyard," "haunted houses." A definite mood of sadness was noted in the associations. With adults this would be expected, but it is worthy of note to find it occurring at such tender years.

Associations here indicating perception of rhythmic relationships were also reported *e.g.* "A long line of men moving very slow," "Of a funeral gradually moving away," "of a long procession of people walking slowly."

(*c*) *Merrymaker's Dance.*—72.4 per cent reported associations, most of which showed origin not only in emotional states corresponding to the stimulus but indicating perceptual activity too. Perception of rhythm was variously translated through associations with the Cinema, Literature, and Home.

(*d*) *Spring Song.*—This was the only record with which about a third of the boys were familiar. Seventy-one per cent reported associa-

tions. The stimulus had a markedly quieting effect indicated in a variety of semi-poetic associations—"lilies swaying gently in a valley," "like when Jesus calmed the sea," "the stately Spanish galleon dipping through the tropics." Perception of the smooth flowing rhythm was indicated in most associations. Of the secondary sexual reaction through cinema associations more will be said later.

(*e*) *Death and the Maiden.*—Seventy per cent of subjects reported associations. They indicated a uniform mood of sadness and usually contained the words "funeral," "burial," and less frequently "church," "grave." Associations were nearly all evoked from a corresponding emotional state. The perceptive type of association very rarely occurred. This result is of interest when we remember that the main musical character of the passage is its harmonies, and, as has been well established, harmonic perception is a later development than either rhythmic or melodic. Cinema associations were frequent—"funeral of some big man on the films," "funeral of the assassinated King," "a funeral procession in foreign countries at the pictures"—are typical associations.

(*f*) *Storm Passage.*—This record compares interestingly with "Steel Foundry" from the associative point of view. Both gave rise to emotional states of disgust and conative attempts to escape by wishing "to smash the record." Unlike the associations aroused by the "Steel Foundry," the "Storm" passage evoked uniformly relevant associations without any word stimulus (*e.g.* the title) being given. A half dozen typical associations were, "a gale storm at sea," "a windy night," "it reminded me of thunder," "like the roar of the ocean in a squall," "a howling gale," "a storm and lightning."

The writer also tried the experiment of having the associations expressed in other than a verbal medium. After the passage had been played, a group of boys was asked to *draw* their thoughts and impressions of what the record was. Whereas the "Steel Foundry" associations, verbal and pictorial, were almost all irrelevant and indicative of wandering trains of thought, the "Storm" passage produced an abundance of relevant associations—drawings of fierce storms on sea and land. One of the most striking suggestions of the violence of the storm was in a horizon of broken telegraph poles and tangled wires; another had a diagonal streak of lightning which represented the various peals of thunder—perception of the rhythmic strips—which are communicated to the listener.

The link between enjoyment, arousal of the conative factor, and of associations, may be represented thus:

Record	Order of enjoyment	Per cent	Order of conative effect	Per cent	Order of associational effect	Per cent
Merrymaker's Dance	1	88.5	1	73.5	3	72.4
Spring Song	2	86	2	60	4	71
Storm Passage	3	60	5	38.6	6	70
Steel Foundry	4	57.6	3	29	1	89
"Death and the Maiden"	5	36	6	25	5	66.8
Asa's Death	6	29	4	8	2	87

The record most enjoyed (Merrymaker's Dance) had almost equal conative and associative effects. The record with the highest associative effects (Steel Foundry) revealed on examination that very few of the associations were relevant to the stimulus. The associations were the products of free trains of thought, associations built for themselves, not at all responsible for the aesthetic enjoyment of the music.

CINEMA ASSOCIATIONS

Today the cinema, with its "talkies" and sound films, is a potent aural factor in the lives of children. When the subjects of this experiment were presented with unknown stimuli there was a marked tendency to interpret them in the light of cinema experience. Four features of these cinema associations were outstanding:

1. Frequency of occurrence.
2. Definiteness.
3. Cinema associations could determine mood.
4. Sexual reactions.

Every record yielded an abundance of cinema associations. A striking feature was the general definiteness of the link. Vague expressions of the type, "it made me think of a film I saw," were exceptions. The majority specified the film or character despite the fact that they had not heard (excluding Spring Song) the records previously. Typical examples were:

Record	Cinema Association
Steel Foundry	King Kong (loudness, noise).
	Tarzan (loudness, noise).
Asa's Death	Barthou's funeral in the newsreel.
	Music played at the pictures when men are sailing down a tropical river.
Merrymaker's Dance	A film called "Mrs. Robinson Crusoe."
	A Mickey Mouse film.

Record	Cinema Association
Spring Song.....	A Silly Symphony—birds dancing on the trees.
	It made me think of a cowboy on his deathbed.
"Death and the Maiden".....	Funeral of the assassinated King.
Storm Passage.....	Of a picture called The Ghoul.
	It made me think of Dracula.

That the cinema associations aroused should be so definite is psychologically important in considering the growth of the child mind. The cinema thus strikes at one of the most primitive tendencies—that of personifying the inanimate. Myers quoting Stout[1] says: "'The cataract or whirlpool appears a living thing to the poet in his poetic moods: For in these moods he ignores the fact that the water is behaving in accordance with certain abstract laws under certain given conditions. This fact is not *ignored* by the savage: It has never been realised by him. Hence what may be called a transcient play of the imagination in the civilized mind is the permanent and serious attitude of the savage mind.' If Stout had at his disposal a more strongly developed character aspect, he would realize that it is the persistence of this so-called mark of the savage that leads to the aesthetic personalization of colours, tones, and music."

The cinema directly emphasises the action of this primitive tendency. Sychronized visual impression strengthens the personalizing of the aural impression, a tendency already deep-rooted in the child mind. This has a restricting effect upon the imagination which the educator has scarcely recognized. Obviously, visual and musical impression should be closely relevant to each other if they are to be used as a combined educative media. But the nature of the cinema associations quoted by the subjects reveals that the visual impression frequently bears no relation to the inherent musical qualities of the stimulus. Thus several subjects reported the "Spring Song" as sad, and on enquiry, typical answers were, "it made me think of a cowboy who was on his deathbed," "of a picture of a man and wife meeting after a long time." This is far removed from the Wagnerian opera where musical communication and visual impression are inherently linked. In the cinema, music is slave of visual impression. When subjects are later presented with unknown stimuli, they interpret them in the experience of the cinema. The tendency to personalize the music is strengthened till the personalization becomes all-important, to the

[1] Myers: "Individual Differences in Listening to Music." *British Journal of Psychology*, Vol. VII, 1914, pp. 68–111.

detriment of the cognitive aspect integrated in a full musical response. The cinema strengthens the primitive tendency to personalize the inanimate.

"The influence of sexual experience and juvenile imagination is shown in the fact that the commonest figures evoked in the recorded associations were those of lovers, dancers, soldiers, villagers, savages, fairies, fauns and goblins."[1] Of the twelve hundred sixty subjects of the writer's enquiry this was equally true, but it is interesting to note that sexual associations, representing a secondary reaction, showed a definite link with the cinema. A marked instance was of a boy of fourteen plus years who interpreted the whole questionnaire to the "Spring Song" in terms of sexual emotion. He associated the stimulus with a film of "two romancing young lovers" and conatively this led to an impulse "to cuddle a girl." In this instance the associative reaction determined the whole affective experience. The "sadness" of two records was frequently identified with the "love picture with a sad ending." Waltzes, moonlight nights, gondolas, carnivals, were also linked with the same source. To the educator, this has considerable significance, because it leads away from the cognitive aspect of musical response, away from the perception of the objective factors of musical structure, and focusses entirely on the affective experience. It suggests a new view on the much debated topic of "musical appreciation" where a stumbling block for many teachers is—should a composition, "programme" or "pure," be introduced to young listeners by a story? Whether the educator uses a relevant story or not, it seems that the average, cinema-fed child of today will interpret the new composition with the aid of his cinema experience. Rarely is this relevant; oftenest it is worthless. In these circumstances there are grounds for the usage of a tale by a skilful teacher at least in the introductory stages of "musical appreciation."

CONCLUSIONS

1. Associations in a musical response are one aspect of the affective experience. The tendency to form associations is widespread even at the early years of ten to fourteen plus.

2. Associations of three types were detected, the first two being musically relevant. Thus associations (*a*) had their origin in a corresponding emotional quality evoked by the musical composition; (*b*) had their origin in a perceptual activity *i.e.*, the association becomes

[1] Myers: *Loc. cit.*

the verbal expression of perception of the elements of musical structure such as rhythm; (*c*) had their origin in wandering thoughts when attention and interest in the stimulus had dwindled. These latter are not of importance musically as they contribute nothing to the aesthetic enjoyment of the composition.

3. Musical enjoyment and high associative power of stimulus are not closely connected, but musical enjoyment has a definite link with associative power plus a strong conative effect of stimulus. The most enjoyed record produced a conative and an associative result of almost equal power.

4. The cinema is a powerful aural factor in the affective development of the modern child. This concerns the musical educator. The cinema makes a strong appeal to the primitive tendency to personalize the inanimate. Musical stimuli become markedly personalized. In reporting cinema associations there is a striking definiteness of association even though the musical stimulus is unknown.

THE PROBLEM OF PRINCIPAL COMPONENTS: II—THE ARGUMENT FOR COMMUNALITIES

C. E. KELLOGG

McGill University

In an earlier article, ("The Problem of Principal Components:—Derivation of Hotelling's Method from Thurstone's," *J. Ed. Psych.* 1936), the writer endeavored to further the systematization of techniques in factor analysis, by clearing up a misunderstanding of the Hotelling method of principal components apparent in Chapter IV of Professor Thurstone's recent treatise (*The Vectors of Mind*, University of Chicago Press, 1935). That article showed that, although Hotelling, in his exposition of the method (*J. Ed. Psych.*, 1933) did not treat of specific factors, and used either unity in the diagonal cells with correlations corrected for attenuation, or reliabilities with raw correlations, his proof, which was restricted to the case of equal number of tests and factors, implied the use of diagonal entries based only upon common factors. It was then shown, by a test with Thurstone's artificial problem given in the chapter cited, and then by a general proof, that the Hotelling method, yielding the principal components directly from the correlation matrix, is valid with communalities and for matrices of any rank.

Since the first article was written there has come to hand Truman L. Kelley's new book (*Essential Traits of Mental Life*, Harvard University Press, 1935), presenting a new technique, and, in the course of a comparison of methods in Chapter III, opposing the use of communalities. He states two points, the first of which involves a misinterpretation of Thurstone's procedure. After reproducing the Thurstone notation for common, specific, and chance portions of test variance, Kelley insists there is no justification for omitting the specific variance from the analysis, since it has just as much reality as any other part of the variance. "It seems to (him) unsound to rule out of the analytical picture which it is sought to portray, by the components derived, any real, that is non-chance, variance," etc. It seems strange to encounter an argument that explicit recognition of specific factors leaves them out of the picture, and that concealing them in the diagonal values and thereby distorting the common factor weights is the proper way to portray them.

Kelley's second point is that the use of communalities obscures relationships, since they are not given, can only be estimated, and so standard errors cannot be calculated, etc. This standpoint is in marked opposition to his own earlier procedure (cf. *Crossroads in the Mind of Man*, Stanford University Press, 1928), which dealt with various combinations of common and specific factors. With the new method he assumes not merely that all the variances should be accounted for, but that this should be done entirely in terms of common factors, represented by orthogonal axes so placed that the new component scores will be uncorrelated.

He suggests using variances in the diagonal cells, covariances (product-moments) in the other cells. This makes the factor weights obtained apply directly to raw test scores, but is not essential to use of the method. Strictly speaking, his suggestion is not quite consistent with Hotelling's procedures mentioned above, which are equally workable with Kelley's technique, and agree with his plan in making no allowance for specific factors, but do allow for the existence of errors of measurement. If correlations are lowered by errors, so are the portions of the variances that can be referred to common factors. So for Kelley's proposal, which corresponds to use of unity in diagonals with raw correlations, we should substitute either variances with covariances corrected for attenuation, or reliability product-moments with those from raw correlations. Allowance will naturally be made for the plan used, in adapting the factor weights for use with test scores.

Without going into unnecessary detail, it may be said that Kelley's new method—which transforms the correlation matrix in one plane at a time, shifting the variances towards one or more of the diagonal cells, while clearing the remaining ones along with the covariances, as far as the rank of the matrix permits—depends upon holding the total variance rigorously constant. Estimates of the stage of convergence towards the proper factor weights are made by working out the correlations from the transformed matrix, by dividing each residual covariance by the geometric mean of the corresponding diagonal cell values. The process is continued until a critical ratio based upon these correlations is reduced below the chance value. Thus a preference for diagonal entries definitely given by the data is natural. But from the Thurstone standpoint, based upon Kelley's earlier teachings, this amounts to an insistence on obtaining factor weights which do not truly represent the relationships among the variables, in order that one may be able to estimate, in a roundabout way, how nearly another set of

untrue weights, obtained in a new trial, would resemble the present ones.

In treating his centroid method (cf. Chapter III, *opus cit.*), Thurstone shows the effect of using various sets of diagonal values with artificial problems, but in handling research data uses only approximations to the communalities. The present article is intended to clinch the argument for communalities, with the method of principal components, by making this comparison in case of a real problem.

For this problem we turn to the excellent thesis of John C. Flanagan (*Factor Analysis in the Study of Personality*, Stanford University Press, 1935), a study supervised by Professor Kelley. In view of Kelley's insistence on inclusion of all the variance in the analysis, Flanagan's results with his second problem, in his Chapter III, seem to have rather marked shortcomings, as the percentages of the total variance (reliabilities in this case) given by the three factors obtained are 48.9, 14.9, and 11.8, respectively, and 24.4 per cent remains unaccounted for. The author simply assumes, following Hotelling, that there are as many common factors as tests, but that the remaining ones would be so small as to be unreliable. The problem evidently invites further study.

We reproduce the correlation matrix in question, giving the intercorrelations between the scores of ten hundred forty-six Bucknell College sophomores on ten different achievement tests (after Brigham), with reliabilities in the diagonal cells: (Flanagan, *opus cit.*, p. 35).

TABLE I

	1	2	3	4	5	6	7	8	9	10
1. Spelling	*.932*	.564	.621	.476	.394	−.022	.044	.389	.344	.328
2. Grammar	.564	*.904*	.742	.577	.472	.013	.158	.429	.426	.383
3. Punctuation	.621	.742	*.907*	.503	.461	.014	.102	.411	.407	.339
4. Vocabulary	.476	.577	.503	*.966*	.688	.030	.334	.548	.494	.546
5. Literature	.394	.472	.461	.688	*.961*	.002	.202	.639	.541	.574
6. Mathematics	−.022	.013	.014	.030	.002	*.986*	.430	−.035	−.075	−.012
7. General Science	.044	.158	.102	.334	.202	.430	*.943*	.258	.183	.276
8. Foreign Literature	.389	.429	.411	.548	.639	−.035	.258	*.915*	.691	.646
9. Fine Arts	.344	.426	.407	.494	.541	−.075	.183	.691	*.920*	.589
10. History	.328	.383	.339	.546	.574	−.012	.276	.646	.589	*.948*

The high reliabilities throughout, and many low correlations, strongly suggest the reality of specific factors.

This problem was worked out by the Hotelling technique, using as a compromise set of diagonal values the geometric means between

TABLE II

(*a*) Reliabilities in Diagonal (Flanagan, p. 37)

Factor	I	II	III	Sum of squares
Spelling	.650	−.249	.432	.672
Grammar	.740	−.144	.409	.736
Punctuation	.716	−.198	.468	.772
Vocabulary	.815	.083	.015	.670
Literature	.795	.014	−.183	.666
Mathematics	.023	.814	.359	.792
General Science	.330	.772	−.000	.704
Foreign Literature	.786	.030	−.337	.733
Fine Arts	.736	−.046	−.337	.657
History	.734	.097	−.381	.694
Sum of squares	4.588	1.399	1.108	

(*b*) Geometric Means in Diagonal

Spelling	.637	−.322	.298	.598
Grammar	.739	−.238	.355	.729
Punctuation	.714	−.309	.378	.748
Vocabulary	.800	.094	.044	.651
Literature	.782	.071	−.165	.644
Mathematics	.019	.601	.461	.574
General Science	.310	.655	.207	.568
Foreign Literature	.777	.119	−.300	.708
Fine Arts	.726	.039	−.337	.642
History	.718	.188	−.313	.649
Sum of squares	4.455	1.111	.945	

(*c*) Largest Correlation or Residual in Diagonal

Spelling	.625	−.325	.180	.5296
Grammar	.736	−.296	.292	.7146
Punctuation	.712	−.371	.319	.7463
Vocabulary	.786	.092	.071	.6313
Literature	.769	.106	−.138	.6216
Mathematics	.018	.421	.456	.3855
General Science	.298	.570	.378	.5566
Foreign Literature	.768	.192	−.285	.7079
Fine Arts	.718	.108	−.292	.6125
History	.704	.245	−.245	.6157
Sum of squares	4.341	.968	.819	

the reliabilities and the highest intercorrelations for each test; namely, *.760, .819, .820, .815, .651, .636, .795, .797,* and *.782.* It was then reworked, using as diagonal values the highest intercorrelation in each column for the first factor, and the highest residual in each column for the later factors, the plan followed by Thurstone with the centroid method. The two solutions will be compared with Flanagan's. We list the factor weights and obtained communalities, and the total contributions to the variances by each factor.

The changes in weights resulting from changes in the diagonals are striking enough. This might be held to reinforce Kelley's argument against statistics "entered by fiat." But we turn next to a comparative test of the weights as a means of reproducing the correlations, *omitting the diagonals* as irrelevant to the problem. They have played their part in yielding the weights, and the reliabilities, whether used in the diagonals or not, have really been taken into account, since the correlations are dependent upon the reliability of tests used, and estimates of communalities have reference to the correlations. Flanagan's factor correlations were calculated and his residuals tabulated for comparison with those from our solutions. Considering only the forty-five different intercorrelations, the results are as follows:

TABLE III

Diagonals	Range of residuals	Sum of absolute values
Reliabilities	−.206.... .097	2.062
Geometric means	−.090.... .223	1.781
Largest correlations	−.062.... .084	.972

As the correlations in the given matrix vary greatly in reliability, a more rigorous check is given by comparison of each residual with the probable error of the corresponding correlation. This was done for the

TABLE IV.—PE RESIDUALS

Range	Number more than four times PE	Absolute sum	Mean	Algebraic sum	Mean
(*a*) −11.839.. 6.291.....	10	125.559	2.790	−36.835	−.8185
(*c*) − 4.105.. 7.5.....	2	60.892	1.353	1.906	.042

first and third of the solutions, the data concerning the second given in Tables II and III being sufficient to show the continuity of the approach to accuracy.

While the third solution is not wholly satisfactory, having two significant residuals out of forty-five, there can be no doubt that it is a great improvement over Flanagan's, which has ten. A fourth factor would surely leave no significant residuals, but there is little reason to suppose that it would stand for anything distinct objectively. In view of the trend of the three solutions compared, it is far more probable that three factors are sufficient, but that a better estimate of the communalities is called for. Probably a repetition of the analysis, using in the diagonals the communalities obtained in this solution, would give a satisfactory result. This being a practical problem, any attempt to eliminate residuals is of course unwarranted, but their reduction to approximately chance values remains the final check.

Flanagan, though he does not explicitly mention it, clearly has in mind Hotelling's test for distinctiveness between factors (cf. the latter's section 6), as a basis for deciding how far to carry the analysis. In such a case as the present, in which the communalities are comparatively low, the solution based on reliabilities does not fulfill either the purpose of accounting for the variances, or that of reproducing the correlations, within reasonable limits, with a small number of factors. Kelley's new method, like Hotelling's, deals with the multi-dimensional distribution of test scores—generalized from the familiar scatter-diagram for zero order coefficients—but keeps track of all the variance at each step, until all significant correlations are eliminated. The final values in the diagonal cells give the roots of the characteristic equation, and show the relative importance of the factors in contributing towards the variances of the system. Kelley's criterion for deciding how far to carry the rotations refers only to the correlations in the transformed matrix, not to those given by the data of the problem. It thus shows how nearly complete, for *whatever diagonal values are used*, the analysis is. Kelley, in the passage cited earlier, qualifies his criticism of Thurstone by saying that his center of gravity method does not demand the use of communalities. We may now return the compliment, and say that Kelley's new method does not demand the use of total variances. He himself says that his method gives the same results as Hotelling's, and, by way of acquiring familiarity with the new technique, the writer has just completed a successful try-out on the same problem from Thurstone treated in the previous article, and using communali-

ties. This is an artificial problem, in which the exact communalities are known. There is no reason to expect any difficulty in use of the method with estimated communalities in a practical problem, except that it should be noted that any marked underestimation will lead to negative values in some of the diagonal cells in the later stages of the analysis, and so will yield some imaginary factors, as Thurstone has shown is the case with the centroid method.

To return to Flanagan's problem—the practical importance of a close approximation to the given correlations will be clear if it conduces to an easier and more convincing interpretation of the factors resulting. Flanagan, to summarize him crudely, interprets his three factors as (1) verbal; (2) mathematics and science—probably numerical; and (3) a difference between practical and cultural information. Plotting the projection of his test vectors on the plane of the second and third factors and keeping in mind the values of the first factor shows that his interpretation is fairly plausible, but while variable 7, General Science, projects along the second axis, variable 6, Mathematics, diverges considerably, so as to have little correlation with either the "practical" or the "cultural" cluster. Plotting similarly our final solution swings all the vectors somewhat clockwise, and Science more than the rest, so that it comes closer to Mathematics. A 45° rotation of the axes is needed to bring them into relation to the vector clusters. This was carried out graphically, giving the following set of factor weights:

TABLE V

Factor	I	II′	III′
Spelling	.625	−.102	.357
Grammar	.736	−.001	.416
Punctuation	.712	−.036	.488
Vocabulary	.786	.114	−.015
Literature	.769	−.022	−.172
Mathematics	.018	.620	.025
General Science	.298	.671	−.133
Foreign Literature	.768	−.064	−.348
Fine Arts	.718	−.131	−.282
History	.704	.000	−.348

The new weights favor Flanagan's interpretation much more strongly than do his own. But the state of affairs in the I–III′ plane is not very satisfactory. Only variable four, Vocabulary, is near the

I axis in that plane, while there are two fairly suggestive clusters, one on each side of that axis. So a rotation to make both these clusters positive, and a slight shift in the I′–II′ plane to approach the clusters a little more closely and also make the interpretation more in line with the given correlations, were carried out similarly, and the final position checked for substantial equality of communalities. Both the second and final sets of weights are given, so that the reader may check the work readily:

TABLE VI

Factor	I′	II′	III″
Spelling	.191	−.102	.694
Grammar	.224	−.001	.813
Punctuation	.160	−.036	.850
Vocabulary	.568	.115	.544
Literature	.664	−.022	.418
Mathematics	−.007	.620	.032
General Science	.306	.671	.114
Foreign Literature	.788	−.064	.290
Fine Arts	.705	−.131	.302
History	.745	.000	.250

Factor	I″	II″	III″	Sum of squares
Spelling	.201	−.083	.694	.529
Grammar	.223	.022	.813	.711
Punctuation	.163	−.017	.850	.749
Vocabulary	.552	.175	.544	.631
Literature	.665	.044	.418	.619
Mathematics	−.067	.618	.032	.387
General Science	.240	.698	.114	.558
Foreign Literature	.788	.016	.290	.705
Fine Arts	.717	−.061	.302	.609
History	.741	.077	.250	.618

A natural interpretation of these factors seems to be: I″—cultural; II″—mathematics and science—perhaps numerical, perhaps sustained reasoning; III″—linguistic usage. These factors may be taken as representing the common elements in the earlier part of the college curriculum, to the extent that it was covered by the set of examinations in question. The remaining parts of the non-chance variances must for

the present be assigned to special interests and aptitudes in the separate subjects. This last set of factors is offered not as a definite solution, but as the most plausible orthogonal set that the writer could secure by simple methods, and one that would probably be practically suggestive in connection with the purposes for which the series of examinations was originally given.

A final question . . . If, in order to achieve simplicity of psychological interpretation, one is to rotate the system of axes from the position given by the principal components, retaining invariant only the configuration of test vectors, why not use the shorter centroid method, and rotate for the final solution? That is, if the special information to be had from the unique system of principal axes is not required, does the method by which they are obtained have any other merits, beyond its greater profundity, which naturally appeals to the academic mind? In the same chapter cited above, Kelley, comparing the centroid and rotation methods, objects to the former because its axes cannot be found by an orthogonal transformation in the test-score distribution space. But this is only a consequence of the orthogonality of the centroid vectors in the unit-variance factor space. In the latter, all orthogonal sets of vectors yield uncorrelated scores, while any departure from orthogonality produces correlation. The test vectors are thus in general not orthogonal, and the principal axes are those on which the sum of the squares of the test-vector projections is maximized—for one factor, then for two leaving the first fixed, etc. In the test space, the only set of orthogonal axes yielding no correlation is that of the principal components, (determinate except in case of multiple roots in the characteristic equation), the test vectors are all orthogonal, any set of axes other than the principal ones must depart from orthogonality to eliminate correlation, and the principal axes maximize the variances accounted for, for any given number of factors, and under the same conditions as just stated for the factor space.

Our present problem applies to the latter space. Thurstone has shown that the centroid method, carried to a large enough number of factors, reproduces the correlation matrix exactly. The writer's previous article proved that the principal factors obtained from any set of factors which completely account for the matrix are identical with those obtained directly from the matrix. So with an artificial problem, involving no errors and with communalities given, or with a real matrix corrected for attenuation and based on so large a population that sampling errors are negligible, so that rank and communalities

can be determined precisely, and it is justifiable to carry the solution to completion, final results will be the same except for calculation errors. But in an ordinary real problem based on a limited experimental population, in which the rank is raised by errors, but only the first few factors can be used, results will differ, because the centroid method places each axis in turn so as to maximize the test-factor correlations, while the principal axes maximize their squares. Accordingly, the latter method favors the tests having the larger communalities, thereby giving greater proportionate influence to the larger and more reliable correlations of the matrix. The PE residuals will thus be smaller, and the results more nearly representative in the long run.

As correlations and residuals are not affected by rotation, it is clear that the centroid and principal components solutions are not equivalent. Whether the increased stability of the latter is worth the extra labor, will naturally depend upon the purposes and preferences of the research worker, as will the choice between the Hotelling and Kelley techniques for obtaining the principal components. In using the Hotelling process, the writer has found it convenient to hasten convergence by taking note of the trend of changes in trial weights, and making larger arbitrary changes accordingly in the next trial, until near the final approximation. A less rough and ready abridgment has now been given by Hotelling himself, in the first number of the new journal, *Psychometrika* ("Simplified Calculation of Principal Components," *Psychometrika*, Vol. I, No. 1, March, 1936, pp. 27–35).

Whatever method of factor analysis is employed, the accuracy of the results will depend upon the degree to which the diagonal entries in the correlation matrix approximate the communalities.

AN EXPERIMENT IN COMPARING RESPONSES TO TWO TYPES OF POETRY, RHYMED AND UNRHYMED, IN NURSERY SCHOOL, KINDERGARTEN, AND FIRST GRADE

DOROTHY BARUCH*

Whittier College

INTRODUCTION

The statement is often made that children prefer rhymed metrical to unrhymed cadenced poetry. The following quotations show such statements of preference for rhymed metrical poems:

What the child loves is the swinging rhyme and jingle. . . . Blank verse which has been written especially for children is highly unsuccessful; they simply don't like it, but nonsense with a definite swing and rhythm, that is a different matter. (P. 190.)[1]

The rhythm appeals to them. . . . The jingle is what attracts him. (P. 176.)[2]

The rhyme and rhythm of poetry are as indispensable to the full growth of a child's nature as the rhythmic movement of sap in a growing tree.[3]

Poems for the youngest child should be chosen for their form as well as for their content. It is the rhyme, the marked rhythm and the alliteration which charm. (P. 77.)[4]

The earliest poetry should be that in which the music is very prominent and the idea absent or not prominent, . . . use and encourage strongly emphasized rhythm in reading poetry, especially in the early work. (P. 370.)[5]

Commentators . . . agree in stressing the appeal made by rhythm, the jingle, the emphatic meter . . . in poetry. (P. 21.)[5]

---

* Assisted by the following students of Broadoaks School of Education, Whittier College: Ethel Alexander, Edith Gemmell, Margaret Ewart, Helen Gillette, Olly Goetz, Margaret Hay, Margaret Henry, Esther Newcomb, Evelyn Walberg, and Caroline Polhemus.

[1] Foster, Josephine: *Busy Childhood.* Appleton, N. Y., 1933.

[2] Field, W. T.: *A Guide to Literature for Children.* Ginn, Boston, 1928.

[3] Penney, Foster, Mix: *Teaching Children's Literature.* Globe Book Co., N. Y., 1934.

[4] Alschuler, R. H.: *Two to Six—Suggestions for Parents of Young Children.* Wm. Morrow Co., N. Y., 1933.

[5] Curry and Clippinger: *Children's Literature.* Rand McNally Co., Chicago, 1920.

Such verses (well chosen nursery-rhymes) seem to be liked, because they are short, are usually dramatic, express considerable action and have a definite rhyme and rhythm. (P. 108.)[1]

Children need the simple striking rhythm of verse, of verse whose rhythm is quite unmistakable. Indeed it is profitable in the first verses that children learn to have an emphatic meter, so that the musical intention may not be missed. (P. 194.)[2]

Most children are pleased with the additional music of rhyme. This is true of all kinds of rhyme, but of course it is the regular terminal rhyme that most children notice and enjoy and remember. (P. 195.)[2]

When one stops at other considerations in relation to young children's reaction to poetry, one commences to wonder whether this statement as to their preference for rhyme and meter is true. One of the chief considerations which makes one pause is the fact that young children when they are left free to create their own poetry spontaneously do so in unrhymed cadenced form. Hughes Mearns and others have pointed this out. In speaking of young children's creative expression, Mearns says:

Native language is often stilted by the adult native belief that all verse should rhyme. (P. 112.)[3]

And again,

To most teachers of young children, to write a poem means to indite a jingle. Without knowing it they step in at the start to block free expression. (P. 127.)[3]

Annie E. Moore remarks similarly:

There is little doubt . . . The necessity to find rhymes becomes a real hindrance to the expression of thought and mood with young children. (P. 174.)[4]

Likewise Lucia B. Morse says:

Free poetic expression of very young children seldom has rhyme, and most frequently partakes of the form of a song or chant with many repeated words. (P. 51.)[5]

---

[1] Foster, J. and Mattson, M.: *Nursery School Procedure.* Appleton, N. Y.

[2] MacClintock, P.: *Literature in the Elementary School.* Univ. of Chicago Press, Chicago, 1928.

[3] *Creative Power.* Doubleday Doran, N. Y., 1930.

[4] *The Primary School.* Houghton Mifflin, Boston, 1925.

[5] "Young Pegasis." *Progressive Education*, March, 1928.

Yet, in the past we have very infrequently, if at all, offered to children poetry of their own kind; namely, cadenced and unrhymed verse. Instead we have taken it for granted that they prefer that which is rhymed and evenly metrical, and have limited their diet to that sort of verse. Have we done wisely in not including the sort of poetry which they naturally make? Have we perhaps not been saying to them subtly by limiting the poetry we give them to rhymed verse, "Your kind isn't good enough for us to care to give you?" Have we not perhaps in this way been just as subtly limiting their expression as they grow old enough to gather the implication that their kind is not good enough? And have we not very much been taking for granted the idea that they markedly prefer verse that possesses meter and rhyme? Do they greatly prefer rhymed verse after all?

The present experiment represents an attempt to test out the last question in certain groups.

PURPOSE

The purpose is to discover whether marked preferences exist for rhymed metrical over unrhymed cadenced verse or *vice versa*, as shown by the responses to readings of both to children in the nursery school, kindergarten and first grade.

SELECTION OF MATERIAL

The first step was to find verse of both rhymed and unrhymed forms, that had been used successfully with children. In consequence, verses were chosen which were reported by teachers to have been liked by past groups of children, but which were, as far as could be determined, new to the groups used in the experiment.

The next step was to eliminate in so far as possible differences in content of the two forms given, or, in other words, to pair poems so that in each pair the same content would appear but in two different forms. Obviously this was necessary or one or the other of the two would undoubtedly have been chosen because of preference for content rather than for form. The only way this seemed feasible was to have poems selected of both types and to write a counterpart of the other type for each poem chosen. Thus each rhymed poem selected was analyzed for content and the content then patterned into an unrhymed version or "counterpart." In the same way for each unrhymed poem a rhymed counterpart was formulated. Three students undertook the writing of the counterparts. Each counterpart was then checked as to similarity

of content by three judges (the two other students and the class instructor taking part in the judging of counterparts). Changes were made until there was agreement that the content was similar in both forms. In presentation care was taken alternately to present first a poem of one form, and next a poem of the other form, to eliminate the possibility of having the children show more favorable responses to a poem because of its having been heard first or last.

The two following sets of poems are given to illustrate the way in which counterparts were arranged to pair with each poem in its original version.

"RAIN," ELSA EISGRUBER FROM SPIN TOP SPIN[1]

Original Version.

Rain
Millions of raindrops
Rain, rain
On my little head
Rain on the green grass
My shoes are all wet
Rain, rain, rain

Rhymed Counterpart.

It started to rain
And is raining down yet,
So when I was outside
My feet got all wet.

It rained on my head
And on the green grass,
After millions of raindrops
The shower will pass.

"STORM OUTSIDE MY WINDOW," DOROTHY W. BARUCH (MS)

Original Version.

The storm blows hard.
The branches
Outside my window
Fight in the wind.
They twist and bend
And hit and slap each other.

---

[1] From Elsa Eisgruber: *Spin Top Spin.* By permission of The Macmillan Company, Publishers.

Their finger-like leaves
Scratch and claw.
They whip and clatter
Angrily
As they battle
Against my window glass.

Rhymed Counterpart.

The storm is blowing hard outside my window;
The branches fight and clatter with wind.
They twist and bend and fight and slap each other
As close against my window pane they're pinned.

Their claw-like leaves—like fingers whip and clatter,
And struggle in the wind, and let it pass.
They scratch and claw, are angry as they battle
Pushed by the wind against my window glass.

(A complete list of the poems used appears at the end.)

SUBJECTS

One hundred forty-eight children were used in this experiment. There were

Fifty-four children at Nursery-school level.
Mean age: 3.5
Age range: 2–4 yrs. 10 mo.
Forty-eight children at Kindergarten level.
Mean age: 5.5
Age range: 5–6 yrs.
Forty-six children at First-grade level.
Mean age: 6.8
Age range: 6 yrs. 3 mos.
7 yrs. 4 mos.

Five poems were read twice at all levels. Thus there were ten hundred eighty presentations of poems at nursery-school level, nine hundred sixty at kindergarten level, and nine hundred twenty at first-grade level.

The nursery-school children, from three nursery school groups, were above average in intelligence, the range of IQ's was one hundred one to one hundred seventy with a median IQ of one hundred thirty-six. As a whole the children came from homes of the two highest socio-economic classes as given by the Sims Scale. This was true also of one of the kindergarten groups, containing twelve of the children. In the

other kindergarten group, the children were average in intelligence and came as a whole from homes of the third and fourth socio-economic classes as given by the Sims Scale. The IQ's of the first kindergarten group ranged from eighty-eight to one hundred thirty-nine with a median of one hundred thirteen, while the range for the second group was ninety-four to one hundred thirty-three with a median of one hundred ten.

The children in the first grades, were found also to be from homes in which the majority would fit into class three and four of the Sims, with only nine per cent in class one. The IQ's for the two first-grade groups had a range of eighty-nine to one hundred thirty-six and ninety-eight to one hundred forty-one, respectively, with median IQ's of one hundred nine and one hundred fourteen.

PROCEDURE AND PRESENTATION IN NURSERY SCHOOL AND KINDERGARTEN

In the nursery school and kindergarten the method of recording and scoring responses was taken from an objective method derived and tried out for reliability by Mary Harlan.[1] According to this method movements and facial expressions are checked and verbalizations written into a record form. Imitative gestures, spontaneous gestures, and approach movements constitute the movements indicative of interest, while leaving the group counts as a response signifying a lack of interest. Laughing and smiling at content count as interest responses, laughing and smiling at unrelated material as lack of interest. Eye focus too is taken into account. Verbalization pertaining to the content is taken as indicative of interest; verbalization about unrelated material, as lack of interest. Each response is negatively or positively scored; the negative then being subtracted from the positive to give the final scores to indicate interest shown.

One experimenter read while another observed and recorded responses. Responses were then scored according to the method cited above, so that preferences of one form of the poem over the other showed up. In nursery school and kindergarten three children of similar age were taken at a time into a room where distractions were limited as much as possible. There were two sittings on different days

[1] Harlan, Mary S.: *An Experimental Study in a Technique for Recording Children's Expressive Behavior Responses to Stories.* Unpublished Master's Thesis, Broadoaks School of Education, Whittier College, Pasadena, Cal., July, 1932, p. 82.

for the five poems. The first sitting included two poems, the second the remaining three. Each poem was read twice and the readings were alternated with rhymed and unrhymed versions first to avoid having order of presentation influence choice. The readers were careful to read both forms with equal dramatic appeal and to let no personal bias enter.[1] A few short, introductory remarks were given to establish a feeling of familiarity on the subject of the poem. During the reading and immediately at the end, time was given for verbalization, which was at once recorded. Facial expressions and movements were checked at the middle line and at the next to last line of each poem.

PROCEDURE AND PRESENTATION IN FIRST GRADE

For the first grade a form was evolved on which the children checked their own preferences. Each preference was scored as a point in favor of one form or the other. Points were then added to obtain the final interest score.

When the two experimenters entered the room, the children, busy at other work, were told that they were to have a surprise and to put away their work materials and put pencils on their desks. One experimenter then said: "The surprise is choosing!" All children love to choose and this was successful as a means of evoking a fine interest. The choice, they were told, was between two poems, both of which told the very same story. The illustration was used that it was much the same choice that would occur if one saw a friend in a brown dress or sweater one day and a pink one the next day. The same friend—just a different dress. By this time the interest was at its height and the recording sheets were passed and explained.

As the blanks were being distributed, questions were asked to initiate interest in the content of the first poem about a drum, for instance: "How many people here have drums?"—or "Has anyone played a drum in a rhythm band?" etc.

The poem was then read and re-read carefully. Before reading the other type of the same poem they were reminded again that this was to be the same story in "a different dress." The counterpart was read twice. Then three questions were asked by the experimenter:

1. Which poem did you like better?
2. Which poem would you like me to read again?
3. Which poem would you like to take home to mother?

[1] It is interesting to know that more than two-thirds of the experimenters preferred rhymed poetry for children.

Time was given for the children to check on their blanks as each question was read.

The experimenters reported that the decision with which the children checked for one form or another showed an awareness and a power to differentiate which had not been anticipated. Real enthusiasm and spontaneity were evidenced.

Taking into consideration the normal attention span of first-grade age-level, it was found better to divide the five poems so as to be used at two sittings than all at one time.

If the poem of a pair read was unrhymed, thus making the counterpart a rhymed version, care was taken that the first poem of the next pair was of the opposite type to avoid a choice depending upon sequence. The average number of children present at a reading was twenty-three.

LIMITATIONS OF THE STUDY

Several limitations existed. First, in spite of careful checking to insure the two forms of each poem's having similar content, differences still existed in the mere way of saying the same thing differently. Thus in its rewriting a version may have either lost or gained interest value. In examining responses, however, (see Table III) it can be seen that no greater interest responses occurred predominantly to either the originals or the counterparts.

A second limitation lay in the small number of children in the experiment. A larger number would have given more adequate samplings.

A third limitation in the preschool levels, lay in the fact that with single observers (and with different observers recording) errors may have entered through recordings and through personal biases of the observers, even though effort was made to counteract this.

A fourth limitation existed in the use of different methods of obtaining responses and of scoring between the preschool and the grade levels. It was felt, however, that while observing responses during the reading was the only possible method in the preschool groups, it would not have been valid at the first-grade level, as children by this time, in the situations which were available, had learned to sit still and listen and to withhold overt responses while being read to.

RESPONSE IN NURSERY SCHOOL AND KINDERGARTEN

At the nursery-school level there were seven hundred sixteen responses, indicating interest in the rhymed, against nine hundred

sixteen indicating interest in the unrhymed—a predominence of two hundred two responses in favor of the unrhymed. Thus 56.2 per cent ±.84 of the responses indicated preference for the unrhymed.

At the kindergarten level there were twelve hundred fourteen responses, indicating interest in the rhymed, as against nineteen hundred forty-nine, indicating interest for the unrhymed—a predominence of six hundred three responses in favor of the unrhymed. Thus 61 per cent ±.57 of the responses indicated preference for the unrhymed.

The totalled responses are shown in Tables I and II.

RESPONSE IN FIRST GRADE

In the first grade (see Table III) there were three hundred eighty-seven responses in favor of the rhymed, as against three hundred thirty in favor of the unrhymed. Thus 54 per cent ±1.27 of the responses were in favor of the rhymed verse. Here the predominence is on the side of the rhymed verse by fifty-seven points.

Neither here nor at the lower levels did there appear to be any particular influence on responses made by the order of the poem's presentation. On some, the first readings received greater interest responses; on some of the second readings. (See Table IV.)[1]

TABLE I.—SUMMARY OF RESPONSES IN NURSERY SCHOOL TO RHYMED AND UNRHYMED POEMS WITH SIMILAR CONTENT

Poem	Response to reading			
		Rhymed		Unrhymed
Rain	c	98	o	144
Envy	c	153	o	289
'Spress	o	116	c	181
Little Wiggle Nose	o	214	c	173
The Unfinished Canary	o	135	c	131
Totals		716		918

o—original.
c—counterpart.

[1] In the nursery school, three poems read first had the larger interest responses, and two that were read second. In kindergarten two read first had larger interest responses and three that were read second. In first grade three that were read first, and two that were read second received the larger interest responses, *i.e.*, no influence could be seen to be borne by the fact of the poems being presented first or last in the pair.

The fact that fifty-four per cent of the responses were in favor of the metrical rhymed verse might be interpreted in three ways; namely,

1. That by this time the children have heard more of the rhymed verse and so it is more familiar and authentic as poetry to them, and that, therefore, also a desire has been created to hear more of the rhymed type of verse.

2. That with the present system of school-music-training their rhythmic sense has developed greatly in the direction of even beat during kindergarten and the first few months of first grade.

3. That a true, even though slight, preference for rhyme has begun.

TABLE II.—SUMMARY OF RESPONSES IN KINDERGARTEN TO RHYMED AND UNRHYMED POEMS WITH SIMILAR CONTENT

Poem	Response to reading			
		Rhymed		Unrhymed
The Woodpecker	o	108	c	154
The Engineer	o	56	c	139
Storm Outside My Window	c	125	o	111
Oranges	c	186	o	324
My Kitty	c	168	o	276
Totals		1246		1949

o—original
c—counterpart.

TABLE III.—SUMMARY OF RESPONSES IN FIRST GRADE TO RHYMED AND UNRHYMED POEMS WITH SIMILAR CONTENT

Poem	Response to reading			
		Rhymed		Unrhymed
The Drum	o	83	c	69
Rain in the Night	o	69	c	67
The Moon	c	82	o	62
Little Snail	c	70	o	75
Chickadee	c	83	o	57
Totals		387		330

o—original.
c—counterpart.

TABLE IV.—SUMMARY BY ORDER OF PRESENTATION OF RHYMED AND UNRHYMED POEMS WITH SIMILAR CONTENT

School	Poem	Response to form presented first			Response to form presented last		
Nursery school.....	Rain	144	(u)	(o)	98	(r)	(c)
Nursery school.....	Envy	153	(r)	(c)	289	(u)	(o)
Nursery school.....	'Spress	181	(u)	(c)	116	(r)	(o)
Nursery school.....	Little Wiggle Nose	214	(r)	(o)	173	(u)	(c)
Nursery school.....	The Unfinished Canary	131	(u)	(c)	135	(r)	(o)
Kindergarten.....	Storm Outside My Window	111	(u)	(o)	125	(r)	(c)
Kindergarten.....	Oranges	186	(r)	(c)	324	(u)	(o)
Kindergarten.....	The Engineer	139	(u)	(c)	56	(r)	(o)
Kindergarten.....	My Kitty	168	(r)	(c)	276	(u)	(o)
Kindergarten.....	The Woodpecker	154	(u)	(c)	108	(r)	(o)
First grade.....	Little Snail	75	(u)	(o)	70	(r)	(c)
First grade.....	The Moon Song	82	(r)	(c)	62	(u)	(o)
First grade.....	The Drum	69	(u)	(c)	83	(r)	(o)
First grade.....	Rain in the Night	69	(r)	(o)	67	(u)	(c)
First grade.....	Chickadee	57	(u)	(o)	83	(r)	(c)

u—unrhymed.
r—rhymed.
o—original.
c—counterpart.

SUMMARY AND CONCLUSIONS

Five pairs of poems, each pair with similar content, one of each pair rhymed and with definite meter, and the other unrhymed and with cadenced rhythm, were read to children in nursery school, in kindergarten and in first grade, with the purpose of studying preferences for one form or the other as indicated by the children's responses.

There were one hundred forty-eight children altogether used in the study; fifty-four at the nursery-school level, forty-eight at the kindergarten level and forty-six in the first grade. Five poems were read twice at each level to all the children, so that there were twenty-nine hundred sixty presentations of poems in all; ten hundred eighty at nursery school level, nine hundred sixty at kindergarten level, and nine hundred twenty in the first grade.

In the nursery-school and kindergarten groups responses were recorded on forms according to an objective method which had in a previous study been tried for reliability. Recordings were done by one

experimenter during the time that the poems were read by a second experimenter.

In the first-grade groups, the children checked their choices on blanks, each child being given a blank for each pair of poems as read.

Responses were scored at each level.

In the nursery-school groups 56.2 per cent $\pm$.84 of the responses indicated preference for the *unrhymed* verses; in the kindergarten 61 per cent $\pm$.57 indicated preference for the *unrhymed;* while in the first grade 54 per cent $\pm$1.27 indicated preference for the *rhymed.*

Thus it can be seen that the statement so often made that young children like only, or markedly better, poetry that has rhyme and definite meter, may be seen to be unfounded in the groups in which this study was made.

Both rhymed metrical verse and unrhymed cadenced verse seemed to appeal definitely to the children.

Should this be the case in other groups as well, it would point to the desirability of including in the poetic diet of young children verse of both forms, so that an appreciation and a sense of value of both may flourish.

LIST OF POEMS USED

For Nursery School.

1. "Rain." Elsa Eisgruber, *Spin Top Spin,* Macmillan, 1930.
2. "Envy." Dorothy W. Baruch (Ms).
3. "Spress." Wymond Garthwaite, *Bread an' Jam,* Harper, 1928.
4. "Little Wiggle Nose." W. Garthwaite, *Ibid.*
5. "The Unfinished Canary." Alicia Aspenwall, *Short Poems for Short People,* Dutton, 1929.

For Kindergarten.

1. "Storm Outside My Window." Dorothy W. Baruch (Ms).
2. "Oranges." Dorothy W. Baruch (Ms).
3. "The Engineer." A. A. Milne, *Now we are Six,* Dutton, 1927.
4. "My Kitty." Lucy Sprague Mitchell, *Here and Now Story Book,* Dutton, 1921.
5. "The Woodpecker." E. Madox Roberts, *Under the Tree,* Viking Press, 1930.

For First Grade.

1. "Little Snail." Hilda Conkling, *Poems by a Little Girl,* Stokes, 1931.
2. "The Moon Song." Hilda Conkling, *Ibid.*
3. "The Drum." John Farrar, D. Knippel, *Poems for the young child,* Whitman, 1932.
4. "Rain in the Night." Amelia Burr, *Ring Around,* Macmillan, 1930.
5. "Chickadee." Hilda Conkling, *Ibid.*

THE DEFINITION OF LEARNING

JOHN NOBLE WASHBURNE

Syracuse University

Learning may be described accurately from either of two points of view: that of the observer or that of the learner. Probably most learners will agree that to them learning appears to be an increase, through experience, of problem-solving ability. More specifically, to the learner, *learning is an increase, through experience, of ability to gain goals in spite of obstacles.* Increase in goal-attaining ability may show itself in a number of ways: By an increase in the complexity of the goals which the learner may conceive and strive for, by the number or the difficulty of the obstacles which may be overcome, by decrease in the amount of help needed or effort expended, or in all of these things.

But if a gain or a loss in one respect is compensated for by an equivalent gain or loss in certain other respects, no learning can properly be said to have taken place. Thus if a person overcomes more obstacles with less effort he has learned, but if he overcomes more obstacles with just that much more effort he has not learned. If he overcomes the same number of obstacles with less effort he has learned, but not if he has received a proportionately greater amount of help, and so on.

In short, learning is clearly a matter of proportion—a proportion between factors involved in problem-solving. An increase in problem-solving ability is a change in this proportion. But just what change? To answer this question the proportion must be stated more precisely—learning must be defined on another level. It must be analyzed into its essential factors and the relationship between these factors clearly stated—stated in some sort of definitive formula. What, then, are the essential factors in problem-solving?

In every problem situation we are aware of a goal (g), and of something intervening between that desired state of affairs and the present state of affairs. If the goal is external there are external intervening conditions (o), such as obstacles, distances, or the like; if the goal is an internal state of affairs, such as the answer to a purely intellectual problem, there are internal difficulties or resistances (r) to be overcome. If we make no attempt to solve the problem, our awareness may stop here. But if we try to solve it—to reach the goal—we are aware of three other factors: (1) Effort (e), (2) external help (h) such as cues, implements, etc.; and (3) internal help (m) such as memories, habits,

and native organismic conditions which extend and intensify the actions arising from "*h*." In brief, from the viewpoint of the experiencer—and that is the viewpoint which commands the largest number of factors—every problem situation, every action which appears to be improvable, involves a relationship between six major factors or groups of factors, which we have designated by the letters listed in Table I.

TABLE I.—DEFINITIONS AND VALUES OF FACTORS IN LEARNING FORMULA

Symbol	Definition	Quantitative value
g	*Goal*, an anticipated state of affairs that the anticipator tries to bring about.*	Varies directly with complexity of anticipated situation, and with *e* when all other factors are constant.
o	*Obstacles*, any form of *resistance* (*external to the organism*) involved in goal attainment.	Varies directly with the actual physical resistance offered by distance, impediments, etc..
r	*Resistance*, within the organism itself, including impeding memories, habits, inferences, conflicting goals, and conflicting organismic activities.	Varies directly with the amount of error, or the expenditure of energy beyond the requirements of *g* and *o*.
e	*Effort*, directed and concentrated (as opposed to diffused and multidirectional) energy.	Varies directly with *o* when all other factors are constant, and possibly also with some combination of metabolic and galvanic action of the organism.
h	*Help* (external to the organism) toward goal attainment—cues, implements, favorable environmental organizations, etc.	Varies directly with number of external factors (including cues) necessary for goal attainment when *e*, *g*, and *o* are constant.
m	*Memories* which aid in imaginal extension of experience *toward* the goal, internal help of all sorts including favorable habits and organismic activities.	Varies inversely with *h* and *e* when all other factors are constant.

* Physiologically a goal may be a gap in an organic (or cerebral) equilibrium pattern.[12]

Some of the relationships between these essential, common factors (or groups of factors), are at once evident. Thus if we assign a quantitative value to the goal according to its difficulty, this value will vary not only with the complexity of the goal (*g*) itself, but with the amount and nature of internal and external interference between the present situation and the goal situation. In other words, the difficulty or

energy demand of g increases proportionally not only with the complexity of g, but with the sum of $o + r$. One may be considered the coefficient of the other, and we may write it $g(o + r)$. Likewise, if we assign a quantitative value to e according to the work it does (amount of resistance it overcomes, etc.) this value will vary with the amount and nature of internal and external help received—*i.e.* the work done will be proportional both to the amount of effort made and to the number and kind of cues, implements, memories, etc. used. The fact that effort and help vary proportionally we may write: $e(h + m)$.

Thus we have three factors which determine the task, $g(o + r)$, and factors things that perform it, $e(h + m)$. When they are equal, when $e(h + m) = g(o + r)$, we say "we are equal to the task," or "the goal may be reached." This is our basic formula and we may now define learning on this level of analysis.

Learning is an increase in the value of m/r in the formula

$$\frac{e(h + m)}{g(o + r)} = 1.$$

Let us see if this condensed statement* covers all the observed facts of learning. In the first place we may see in the following formula (which is derived from our basic formula by moving m/r to one side of the equation and all the other factors to the other side) just what an increase in m/r implies—*i.e.* what changes characteristically accompany learning: $\frac{m}{r} = \frac{g}{e}\left(\frac{o}{r} + 1\right) - \frac{h}{r}$. Here it is evident that if the value of m/r increases, (that is, if m increases or r decreases, or both) and all other factors are held constant except h, then h must decrease. In this case we have the phenomenon of learning being proportional to *cue reduction*. Our formula, in other words, corresponds with Hollingworth's description of learning.[3]

If, however, when m/r increases all other factors are held constant except o, then o must increase; and in this case learning is *not* a matter of

---

* We may note, in passing, a resemblance between the learning formula and the formula for a current in a simple circuit including an external resistance R, a cell of electro-motive force E, and internal resistance r, $i = \frac{E}{R + r}$.

If we substitute g for i and o for r, this formula is simply a transposition of the learning formula, with the coefficient of e left out. The correspondence of E, R, r in this formula to e, o, r, in the learning formula is obvious. So far as the substitution of g for i is concerned, the writer has shown elsewhere[13] the relation between goal and the completion of an electro-chemical circuit in the cerebral cortex.

cue reduction, but a matter of the number of obstacles which may be overcome. Thus the formula includes Koffka's point that learning is seldom or never altogether specific and involves "*a change in ability*" to reach an objective in spite of even novel obstacles.[6] Moreover, Koffka's point that *learning involves goal elaboration* is covered by the fact that if all factors but g are held constant, then an increase in m/r must involve an increase of g. Similarly, one can see that *learning is equivalent to effort reduction*, if, when we increase m/r, we hold all other factors constant except e.

Finally, we may turn from the discussion of the equivalents of m/r to a direct account of these two factors. In the first place, it appears from the second formula that all learning, whether cue reduction, effort reduction, or goal increase, *i.e.* any increase in the total value of the right-hand side of the equation, is equivalent to either an increase in m, or a decrease in r, or both, on the left-hand side of the equation. In the one case (an increase in m) we have the type of learning which consists largely in *elaboration*—the forming of new associations and imaginal extensions as a result of exploration. In the other case (a decrease in r) we have the type of learning which consists in *simplification*—cue reduction, telescoping, substitution, and automatization as a result of drill.

These two general types of learning (elaboration and simplification) are clearly illustrated in the case of an individual overcoming somewhat novel physical obstacles on the way to a geographic goal. His first experience consists almost wholly in exploring, *i.e.*, in elaborating his concept of how the goal may be reached—in short, of building up m. This may involve an increase in effort (e) proportional to the increase in real and inferred obstacles (o and r) minus whatever help may be obtained. It also involves an increase in h. Both sides of the equation $e(h + m) = g(o + r)$ increase, but the end result is that they are equal. A *unit* is established—a unit by which learning may be measured. This first gaining of the goal, or "initial achievement,"[7] is prerequisite to the discernment or evaluation of learning. For without knowledge of the difficulty of initial achievement it is impossible to say how much easier are subsequent achievements—in other words, how much has been learned. After "achievement" or *the unit* $\frac{e(h + m)}{g(o + r)} = 1$ has become established the process of simplification becomes predominant.

The explorer in our example finds short cuts and other means of saving himself effort in subsequent journeys. By substituting memory and helpful inferences (m) for actual perceptions (h) he saves himself steps; by eliminating conflicting suggestions he reduces r. Finally, by breaking down his goal into way points—into a quasi reflexive series of focalizations—he automatizes his journey; that is, he reduces the conscious effort involved to a minimum.

To summarize the foregoing account of learning, it consists in:

1. Initial achievement, *i.e.* the establishment of the ratio $e(h + m) = g(o + r)$.
2. An increase in the value m/r in this ratio.

If initial achievement, or insight, is not spontaneously and immediately present but is accomplished by increasing the value of $e(h + m)$ while holding $g(o + r)$ constant the behavior is describable as trial and error. In the reverse case the behavior resembles that which is observed in conditioned reflex experiments. The bearing of the learning formula on Trial and Error and the conditioned reflex will be further discussed in a later article, "*The Learning Ratio and its Application.*" In that article also the attempt will be made to show that the formula, even when precise quantitative values cannot be assigned to its factors, may be employed in the making of useful predictions.

Our present concern is only with the question of whether the formula and the verbal statement which gave rise to it meet the basic requirements of a definition; namely, that they include all those factors (in this case changes in behavior) which may properly be classed under the defined word (learning) and exclude all those which may not. These requirements, we have attempted to show, have been met.

Many definitions of learning do not meet them—some because they include too little and some because they include too much. Among those statements which include too little is Hollingworth's "learning is cue reduction."[3] This leaves out of account the learning in which cues are not reduced but effort is, the learning in which cues are increased, substituted, and combined, and the learning in which the cues remain the same but the behavior changes.[11] Likewise Thorndike's "learning is connection"[10] (and connecting is a statistical probability that if a given situation occurs a given response will occur),[9] leaves out of account the learning which is cue reduction and the learning which is effort reduction—where the probability of the occurrence of a given response in a given situation remains the same, but where the

effort and help employed are reduced. Among those statements which include too much are Woodworth's* "learning amounts to a change or modification of behavior."[15] Also Hunter's "learning is taking place wherever behavior shows a progressive change or trend, with a repetition of the same stimulating situation and where the change cannot be accounted for on the basis of fatigue or of receptor and effector change."[5]

Strictly applied, these definitions mean that changes in behavior which arise from forgetting, confusion, and retroactive inhibition must also be classed as learning. How unacceptable such a notion of learning is may be seen in its practical application in the construction of tests to measure learning achievement. If such tests give credit for all "progressive changes" in behavior which are not due to fatigue, etc., whether these changes are in the direction of precision or confusion, remembering or forgetting, they do not measure learning in any commonly accepted meaning of that term. Likewise tests based upon definitions which are too narrow yield one-sided and misleading results. If habit formation, cue reduction, and stable "connections" are tested for, and rewarded by high marks, all those features of education which give rise to drilled, regimented, and automatic response are over-emphasized, and all those features which give rise to individual, creative, and original response are under-emphasized. Tests based upon the assumption that learning is an improvement in problem-solving ability† will encourage almost equally those educational procedures which foster individual enterprise and variation and those which foster the acquisition of routine and automatic behavior patterns.

Many definitions of learning resemble the one here proposed. The *New Standard Dictionary*,‡ although defining the word quite differently, goes on to explain: "The primary and essential thing about all processes of learning is the establishment of improved systems of bodily and mental reactions to the ever more complex problems which abound in man's environment." The same dictionary defines a problem as "something proposed to be done." In other words, learning is the

---

* A more satisfactory definition of Woodworth's is quoted elsewhere in this article.

† It is only when learning is so defined, that intelligence may properly be said to be "ability to learn." When learning is thought to be something narrower, such as remembering or habit formation, it is relatively, though not altogether, independent of intelligence.

‡ Funk and Wagnals, 1926, N. Y.

establishment of improved reactions to things which are proposed to be done. But proposed by whom? When an end is proposed by an experimentor and is either not comprehended or not accepted by the learner, the situation is entirely different from the one in which the learner has an end of action definitely in his own mind, and strives for it. In the second case the learner has a problem, in the first case none.

Very likely a learner's goal is implied whenever the term "problem-solving" is used. But it needs to be clearly distinguished, for it is altogether omitted from some concepts of learning upon which classical experiments are based, and it is vaguely and uncertainly implied in many others. It is frequently assumed, when it is not stated, that learning is an increased economy in organic energy. Humphrey[4] says "Where there is learning the later actions of the series ordinarily differ from the earlier ones in the direction of the organism's advantage . . . The series will generally show a general approximation to an optimal term, optimal, that is to say, from the point of view of systemic conservation" (p. 105). But the mere saving of energy without achievement, or the mere formation of a habit (as in the case of a tic) without satisfying some desire, without attaining some goal, is by no means all of learning, if it is learning at all.*

Some psychologists argue[3] that what follows cannot be a cause of what goes before, and therefore goal or purpose should not be used as an explanation of, or even as a part of the description of, the behavior that leads to it. To these people it seems correct to speak of behavior as a response to a preceding or present stimulus but not as action toward a goal—or a given state of affairs not yet present. Without taking up the question of causal relationship, it may be stated as an obvious fact that so far as direct personal experience is concerned, goals and purposes are as much a reality as reflexes. We are at least as clear about what it means to seek a goal as we are about what it means to act automatically. So far as the observed behavior of others is concerned, it will probably be conceded that the observation of correlations between behavior and what follows it, namely, the goal, is just as objective as the observation of correlations between behavior and what precedes it,[14] namely, the stimulus.

---

* The statement would be satisfactory if it were revised to read: *Learning is increase, through experience, of organic efficiency;* in which efficiency means (to paraphase the mechanical definition) the ratio between amount of organic energy expended and the amount of work performed; and work means the amount of resistance overcome in the performance of a given task.

Finally, if it is subjective to say that an animal's actions arise from its sense of goal, it is just as subjective to say that its actions arise from a sense of hunger, pain, or any other stimulus. The idea that subjectivity is necessarily unscientific is incorrect. Neither subjectivity nor objectivity is scientific *per se*. All that matters is the *consistency*, *reliability*, and *agreement* with which phenomena are reported. And in this respect goals are as scientifically utilizable as stimuli. Not only is the concept of goal scientifically admissable but it is indispensible. It is present in the broader definitions of learning even when the *terms* "goal," "purpose," "task," or the like, are avoided. Koffka defines learning as "a change in ability,"[6] Woodworth as "better adaptation of the response to the situation."[15] These statements have no meaning unless they refer to the achievement of a purpose. There is no question of ability in random activity, or in reflex activity, no problems involved, no improvement, no criteria of better or worse. To illustrate: Three men, in the same environment, are reacting in different ways; one is standing and looking, one is walking, one is picking things up from the ground. Which man shows greater "ability," better "adaptation," "improved reactions to problems"? The questions are entirely unanswerable unless we know something about, or assume something about, their purposes or goals.

Vague assumptions are disturbing to scientific description. If the concept of goal has a place in our definition, let us make that place as definite as possible. This we have done by restating our definition of learning in the form of an equation showing the precise relationship between the goal factor (g) and the other observed factors. By clarifying the relationship between variables this equation also resolves the apparent conflict between many partial descriptions of learning. Moreover, as will be shown in a subsequent article, it is an aid to prediction and experimentation. But so far as the requirements of a definition are concerned, they are adequately met by the broad, general statement: *Learning is an increase, through experience, of problem-solving ability.*

REFERENCES

1. Dunlap, Knight: *Response Psychology, Psychologies of* 1930. Clark Univ. Press.
2. Hollingworth, H. L.: *Psychology.* N. Y., D. Appleton & Co., 1928.
3. Hollingworth, H. L.: "What Is Learning?" *Sci. Mo.*, 1932.
4. Humphrey, G.: *The Nature of Learning in Relation to the Living System.* London, 1933.

5. Hunter, W. S.: *Experimental Studies of Learning, Foundations of Exper. Psy.*, ed. by C. Murchison. Worcester, Mass., 1929.
6. Koffka, Kurt: *Growth of the Mind.* N. Y., Harcourt Brace, 1928.
7. Köhler, Wolfgang: *Gestalt Psychology.* N. Y., Horace Liveright, 1929.
8. Skinner, B. F.: "The Concept of the Reflex in the Description of Behavior." *Journ. Gen. Psy.*, Vol. IV, 1931, pp. 427–457.
9. Thorndike, E. L.: *Fundamentals of Learning.* N. Y., Teachers College, Bur. of Pub., 1932.
10. Thorndike, E. L.: *Human Learning.* N. Y., The Century Co., 1931.
11. Washburn, Margaret F.: *A System of Motor Psychology, Psychologies of* 1930. Clark Univ. Press, 1930.
12. Washburne, John N.: "An Electro-Chemical Theory of Learning." *Jour. Ed. Psy.*, Vol. XXVI, Feb., 1935, pp. 2, 99–121.
13. Watson, John B.: *Behaviorism.* N. Y., W. W. Norton, 1930.
14. Wheeler, Raymond H.: *The Science of Psychology.* N. Y., Crowell, 1929.
15. Woodworth, Robt. S.: *Psychology.* N. Y., Henry Holt & Co., 1929.

QUALITATIVE WHOLES: CLASSROOM EXPERIMENTS

MAY V. SEAGOE

University of California at Los Angeles

Elsewhere in the literature have appeared articles presenting the hypothesis that the degree of internal organization within the material governs the relative efficiency of whole and part presentation, together with reports of laboratory experiments demonstrating that the hypothesis seems to hold with material largely ideational when the criterion is mastery. It is the purpose of the present experimental series to demonstrate the meaning of the hypothesis for classroom method. Other applications have already been suggested by Decroly in language arts, Barton in typing, and R. Brown in music.

The subjects were drawn from three 8*B* classes in the Garvey School District near Los Angeles. They were paired on the basis of age, intelligence, sex, and classroom. The experiments were carried out by a teacher familiar with objective testing techniques under verbatim instructions.[1]

EXPERIMENT 1. SOCIAL STUDIES MATERIAL

Perhaps the application of the principle of wholeness most widely discussed at present is found in the integration of the social studies. One aspect of this trend is the unification of history and geography. Experiment 1 was devised to test the validity of the assumption that the synthesis of such material results in better learning than does its segregation.

The subjects were thirty-six pairs approximately equivalent in age, sex, intelligence, and classroom background. A slight superiority in age on the part of one group was offset by a slight superiority in intelligence quotient in the other. The difference between mental ages was less than two months.

The material consisted of a two hundred seventy-seven word description of the historical and geographic factors leading to the establishment of a penal colony in Australia. This unit was chosen because the material was foreign to the course of study and very unlikely to

[1] The author is indebted to Ralph Cutlip and to the administration of the Garvey District for coöperation in the entire series of experiments, and to L. M. Terman, J. C. Almack, R. Bell, and E. R. Hilgard, of Stanford University for advice.

have been previously contacted. To the group taught by the whole method the material was presented interwoven in its relationships. To the group taught by the part method the material dealing with the geography of the country was segregated from that dealing with history. The same facts and the same wording except for grammatical changes necessitated by the segregation were used in both forms. According to the hypothesis, the interwoven form would prove superior because it more nearly meets the criteria of wholeness, placing the parts in a relationship which gives them more meaning than they have in isolated form or in juxtaposition.

Each class was divided into two groups according to prescribed lists. The interwoven material in its entirety was distributed to Group W and the segregated historical material to Group P. The children were all instructed to study the material in order to remember as many facts as possible, but not to attempt to memorize. They were told they would have four minutes to study and later in the day would be tested on the facts presented. No talking or note-taking was allowed.

A short period of easy drill on arithmetic fundamentals followed. The interwoven material was then again distributed to Group W and the segregated *geographic* material was given to Group P. The students were instructed to study as before for six minutes. The time limit was based on the relative number of words involved in the two P selections.

Following another short period of work in arithmetic a completion test of nineteen items was given. There was a fifteen-minute time limit. Twenty-eight days later the test was repeated without warning. Of the nineteen questions, seventeen involved isolated facts, one a statement of the relationship, and one a review of the main idea. Correlating odd-even items on the first eighteen questions and raising the result by the Spearman-Brown prophecy formula the obtained reliability of the measure was $.847 \pm .024$.

Results were evaluated in terms of scores on the factual questions supplemented by the number of words written in response to the request for the essential idea of the selection and by a subjective estimate as to whether the relationship between geographic and historical factors was expressed in the statement of the main thought. The results are summarized in Table I. Both score and words written in describing the central thought show some superiority for whole presentation, both in terms of mastery and retention over a period of twenty-eight days. The subjective estimate indicated that in original learning 63.6 per cent

TABLE I.—RESULTS, EXPERIMENT 1

Item	Test score		Words discussion	
	Mastery	Retention	Mastery	Retention
Mean: whole group	19.85	14.33	26.29	18.56
Part group	17.18	13.42	20.68	14.02
PE: whole group	0.48	0.59	1.70	1.84
Part group	0.62	0.62	2.07	1.66
Difference in means	W 2.67	W 0.91	W 5.61	W 4.54
PE of difference	0.78	0.86	2.68	2.48
Ratio	W 3.42	W 1.06	W 2.09	W 1.83
Chances in one hundred true	W 99	W 76	W 92	W 89

of those using the whole method mentioned the relationship between factors as opposed to 15.2 per cent of those using segregated materials. The corresponding figures on retention were 36.4 per cent and 15.2 per cent for the whole and part groups respectively.

The principle of whole presentation of a closely integrated unit seems to apply to the teaching of interwoven social studies material as judged by the consistency of the differences obtained.

EXPERIMENT 2. FRENCH TRANSLATION

Another application of the principle is found in the teaching of foreign languages. There have been numerous studies comparing the introduction of new words in vocabulary form with their introduction in context. Experiment 2 is related to these studies in the same manner that the problem of qualitative wholes is related to those of context and degree of meaningful organization. The present experiment differs in that the comparison is between exactly the same words learned in relationship and in isolation.

The subjects were thirty-two pairs of children approximately equivalent in age, sex, intelligence, and classroom background. Slight inequalities counterbalanced each other as in the case of Experiment 1. None of the children had been exposed to French either at home or at school. Most of the subjects were the same as those used in Experiment 1; for this reason the particular pupils using the whole and part methods were reversed in order to rotate out any significant characteristics not held constant.

The material consisted of a thirty-word joke involving twenty-five different words. The group using the whole method studied the words

written as a joke with the corresponding English translation below each word. The group using the part method learned the English equivalents for the French words listed in alphabetical oder.

Each class was divided into two groups according to prescribed lists. After the material was distributed the class was instructed not to attempt pronunciation or memorization, but to try to learn the English equivalents. The teacher read the French only of each form in turn to satisfy the children's curiosity concerning pronunciation. There was to be no talking, and after twenty minutes the study forms and any study notes were collected.

The test forms were the same as the study forms except that each English word was replaced by a blank. The group which studied the vocabulary form was tested first on the context, then on the vocabulary form. The group which studied the context form reversed this order. Both groups were tested at the same time, and the time limit for each test was eight minutes. The unfamiliar form was given first to minimize rote memorization. The testing procedure was repeated twenty-eight days later without any earlier warning.

The reliability of the context test obtained by correlating odd-even blanks and using the Spearman-Brown formula was .980 ± .003, that of the vocabulary test .904 ± .003, and that of the two tests combined .951 ± .008.

The results are summarized in Table II. There is a tendency for each group to excel the other when tested on the form on which learning occurred, as one would predict. However, the degree of

TABLE II.—RESULTS, EXPERIMENT 2

Item	"Whole" test		"Part" test		Combined score	
	Mastery	Retention	Mastery	Retention	Mastery	Retention
Per cent correct: whole method	93.53	72.07	66.21	31.24	80.00	52.27
Part method	66.87	33.97	77.24	33.24	70.91	32.73
Mean score: Whole	28.06	21.62	16.56	7.81	44.00	28.75
Part	20.06	10.19	19.31	8.31	39.00	18.00
PE of mean: whole	0.40	0.74	0.45	0.36	0.76	1.02
Part	0.65	0.72	0.46	0.40	1.08	1.10
Difference in mean scores	W 8.00	W 11.43	P 2.75	P 0.50	W 5.00	W 10.73
PE of difference	0.76	1.03	0.64	0.54	1.32	1.50
Critical ratio	W 10.53	W 11.10	P 4 .30	P 0.93	W 3.79	W 7.17
Chances in one hundred true	W 100	W 100	P 100	P 74	W 99	W 100

superiority of the context group was greater for its form than that of the vocabulary group resulting in superiority for the former in combined score. The results seem to show that learning in context form approaches equality with learning in vocabulary when the test is for word recognition, but that learning in context is definitely superior when the test is on that content. These tendencies are evident both in original learning and in retention. Another step would be necessary to determine whether the transfer to new content would show a similar relationship.

The principle of whole presentation of a meaningful unit seems to apply to the teaching of French vocabulary.

EXPERIMENT 3. MANUSCRIPT WRITING

A third application of the principle is evident in the teaching of handwriting. Those who advocate drilling on the ovals and strokes which form letters are using a form of part method, while those who stress practice on words are advocating a whole approach.

The subjects were thirty-five pairs of children approximately equivalent in age, sex, intelligence, and classroom background. Slight inequalities again counterbalanced each other.

Because of the fact that motor skills such as handwriting are not very highly correlated with intelligence a rotation technique was added to further assure equivalence of the groups. One group learned the first word by the whole method and the second word by the part method; the other group reversed the order of the methods.

None of the subjects used had previously received instruction in manuscript writing, though all could use cursive script. Two words in manuscript form, "revenue" and "manoeuver" were therefore selected for practice. The whole method consisted of an initial trial, ten practice trials, and a final trial. The part method consisted of an initial trial on the work, practice on each stroke amounting to ten times the frequency of occurrence of that stroke in the word, and a final trial on the word. Each group learned one word by each method. The time limit for practice was sufficient to allow eighty per cent of the class to finish. Without warning twenty-eight days later the students were asked to write the words again in manuscript form, no model being presented.

Scoring was a little difficult. To each sample of writing on pretest, mastery and retention was assigned an identification number. Samples were shuffled in such a way that the rater could not identify whole

from part specimens, or pretest from end-test or retention check. Specimens for each word representing five grades of performance were selected and arranged as a rating scale, "1" being assigned to the highest "2" to the next, etc. Each specimen was then rated three times independently, the average of the three ratings constituting the score for the specimen.

The results are summarized in Table III. That the groups were approximately equivalent in ability at the start is shown in the small difference between means on the pre-test. Any actual difference was

TABLE III.—RESULTS, EXPERIMENT 3

Item	"Revenue"			"Manoeuvre"		
	Pretest	Mastery	Retention	Pretest	Mastery	Retention
Mean: whole method	2.92	2.51	3.20	3.14	2.59	3.33
Part method	3.07	2.98	3.50	3.16	2.90	3.71
PE: whole method	0.10	0.10	0.14	0.10	0.11	0.14
Part method	0.09	0.09	0.16	0.10	0.11	0.13
Difference in means	W 0.15	W 0.47	W 0.30	W 0.02	W 0.31	W 0.38
PE of difference	0.13	0.13	0.21	0.14	0.16	0.19
Critical ratio	W 1.15	W 3.62	W 1.43	W 0.14	W 1.94	W 2.00
Chances in one hundred true	W 78	W 99	W 84	W 54	W 91	W 91

rotated out. The whole method shows a slight but consistent superiority. A comparison in terms of gain over pre-test shows that on the two words the average gain for those using the whole method was .48 points and that for those using the part method .18 points. The loss in retention was approximately the same for the two methods.

The scoring in this experiment was more subjective than in the others, the practice period was comparatively short, and the results from the mirror-drawing experiment reported elsewhere would lead one to expect very little difference. In spite of these factors the principle of whole presentation seems to result in small though consistent differences in the teaching of writing.

EXPERIMENT 4. NONSENSE POEM

Finally, in order to test the conclusions of earlier experimenters from the point of view of a qualitative definition of the term whole,

Ballard's nonsense poem[1] was presented. The poem had the advantage over meaningful material that it was similar to material commonly used in this field from 1900 to 1917, and that it was equally new to all subjects. It had the advantage over a nonsense series that rhythm and rhyme introduced some organization, one element essential in a whole. The poem had the disadvantage that, being devoid of meaning, it formed a very loose configuration and hence any difference obtained would theoretically be smaller. Whatever elements of wholeness were present were matters of form rather than thought.

The subjects were paired in three groups of twenty-three each on the basis of age, sex, intelligence, and classroom background. The groups were equivalent, the estimated mean mental ages being 160.01, 160.77 and 160.51 for the "total," "whole," and "part" groups respectively.

The material consisted of three verses of four lines each, possessing both rhyme and rhythm. In qualitative terms, it is probable the verse is the strongest unit since no unit smaller is complete in both rhythm and rhyme, and larger units are merely sums of smaller units.

Three methods were used. In the part method the subject wrote one line the specified number of times before going to the next line. In the whole method he wrote the first verse a specified number of times before going to the next verse. In the "total" method the subject wrote all three verses before repeating. If the assumption that efficiency in learning varies with the degree of organization of the unit is valid, the expected order of superiority would be whole or verse, total or poem, and part or line.

Each class was divided into three groups according to prescribed lists, the selection read through by the teacher once to satisfy the child's curiosity and the method to be used by each group was indicated with the request that each unit be copied ten times. Papers were collected when eighty per cent of the children had completed their study and each child was asked to reproduce as much of the poem as he could. The time limit was eight minutes. Twenty-eight days later without warning the pupils were asked to recall as much of the material as they could.

In scoring, one point was assigned for each word occurring in the proper sequence within the line, one-half point for words given in list

---

[1] Ballard, P. B.: "Oblivescence and Reminiscence." *Brit. J. Psychol. Monog. Suppl.*, Vol. I, No. 2, p. 7.

form or for correct words incorrectly placed. An error of one letter in spelling was allowed.

The results are summarized in Table IV. Means are in terms of words recalled. Generally speaking, the results substantiate the

TABLE IV.—RESULTS, EXPERIMENT 4

Item	Total method		Whole method		Part method	
	Mastery	Retention	Mastery	Retention	Mastery	Retention
Mean	15.44	5.87	18.57	6.22	13.09	6.48
PE of mean	1.17	0.92	1.40	0.95	1.03	0.92
Groups compared	T-W	T-W	W-P	W-P	T-P	T-P
Difference in means	W 1.60	W 0.18	W 2.59	W 0.22	T 0.99	T 0.04
PE of difference	1.82	1.32	1.74	1.32	1.56	1.30
Critical ratio	W 0.88	W 0.14	W 1.49	W 0.17	T 0.63	T 0.03
Chances in one hundred true	W 73	W 54	W 84	W 54	W 67	W 51

hypothesis although the differences are small. The reason for the small differences probably lies in the nonmeaningful nature of the material. The differences that occur are in the expected directions.

SUMMARY

This series of articles has presented the hypothesis that the criterion governing the relative efficiency of whole and part presentation is a qualitative rather than a quantitative one, has formulated a definition of whole according to the Gestalt school, and has substantiated the hypothesis for certain laboratory and classroom learnings.

These findings are in harmony with earlier studies using comparable material, *e.g.*, those of Barton, Decroly, Roberta Brown, and even that of Mather and Kline. Materials whose classification is uncertain, such as rote memorization of nonsense or meaningful material, mazes, and finger sequences, yield contradictory results.

It must be remembered that the largest unit is not necessarily the whole, that what constitutes a whole for the adult may not be a whole for the child with his limited grasp of meanings, and that what constitutes a whole for one purpose may be a part when the purpose changes. The concept is a relative rather than an absolute one.

The concept has possible implications for several phases of education. (1) In curriculum-making the present emphasis on integration would seem to have psychological justification within limits. Although

the traditional subject or part organizations may lead to some saving of time in immediate reproduction of such parts, the integrated wholes will be more economical in eventual mastery and retention. The selection of life-like units within the grasp of the learner would seem justified. Conversely there would seem to be no merit in merely grouping together unrelated units in the name of correlation unless there is a normal relationship unifying them. (2) In the field of teaching method, the organization of material into meaningful wholes should prove more economical than the teaching of isolated facts. (3) In the field of administration these facts suggest the breaking down of many subject barriers as well as of departmental organization at least on the lower levels and the substitution of a more flexible system wherein materials may be organized around life-like purposes.

THE SPEED FACTOR IN READING PERFORMANCE

VERNA L. ANDERSON AND MILES A. TINKER

University of Minnesota

There is rather general agreement that two important factors in reading performance are speed and comprehension. Examination of the existing reading tests, however, discloses little agreement among authors as to what constitutes a measure of either speed or comprehension.

Tinker[1] has demonstrated that the word "comprehension" or "speed" in the title of a test gives no assurance that the test measures a general comprehension or a general speed factor in reading. Indeed, evidence to the contrary is rather conclusive. Unless tests of "comprehension" are strictly comparable in textural material and kind of response required, they do not yield comparable scores. The same is true for speed of reading tests. Tinker concludes that "there are many reading skills which are somewhat independent, rather than either a general silent reading ability, a general comprehension ability, or a general speed of reading ability."

There have been many attempts to determine the relationship between "speed" and "comprehension" in reading. The customary way of investigating this relationship has been to correlate speed test scores with comprehension test scores. Ordinarily the rate measure has been secured from reading easy narrative prose, and the comprehension measure from answering questions on paragraphs of relatively difficult material.

The claim that such correlations indicate the relationship between speed and comprehension is misleading. *They merely signify the degree of relation between specific speed scores and specific comprehension scores.*

In the practical reading situation the important problem is the amount of relationship between rate of comprehension (speed of reading) and degree of comprehension (comprehension score) in a specific reading situation, *i.e.*, the correlation between rate of reading history and comprehension of the history material, etc. It would seem, therefore, that the only adequate method of discovering the true relationship between rate of work and comprehension in reading is to

---

[1] Tinker, M. A.: "The Relation of Speed to Comprehension in Reading." *School and Society*, Vol. XXXVI, 1932, pp. 158–160.

measure rate and comprehension on the same or strictly comparable material in each specific reading situation. This approach has been employed in the study reported below.

The problem in this experiment is, therefore, to determine the relationship between rate of work ("speed") and achievement scores ("comprehension") in reading when both scores are obtained from the same textual material. Several reading skills, in which content and method of response to the material vary considerably, will be sampled.

The first five parts of the Iowa Silent Reading Test were given to 110 college sophomores. These parts were designed to measure: (1) Paragraph meaning; (2) word meaning; (3) paragraph organization; (4) sentence meaning; and (5) the location of information. The composite of the five tests is called "total comprehension." For the time being we shall assume that the number of items done correctly represents a reading ability score. It is immaterial to us at this time whether such a score is called "comprehension" or by some other name. The reliability coefficients cited for these tests are relatively high; for the "total comprehension" score it is .96.

The tests were given individually and the rate of reading of each subject was timed with a stop-watch. Time limits were set empirically so the fastest student could almost but not quite complete each test.

The subjects were instructed to work rapidly and consistently, but not to sacrifice accuracy for speed. Each subject was allowed to work at a test until the experimentally determined standard time had elapsed, at which point he was interrupted and a line drawn across the page below the last item attempted. Instructions were then given to complete the test, and when the last item was finished the total time required for the whole test was recorded.

Four scores were derived from the data: (1) Score in standard time; (2) number of items attempted in standard time; (3) total time taken to complete the test; and (4) score in unlimited time. The scores in unlimited time did not yield a normal distribution, and, therefore, were not employed in the statistical comparisons.

Correlations were computed to show: (1) To what degree the number of items attempted in standard time is a measure of rate of work; and (2) the relation of the two measures of rate of work to test score in standard time (comprehension). The coefficients of correlation are given in Table I.

It seems reasonable to assume that the total time taken to do all the items on a test is a satisfactory measure of *rate of work* since it actually

TABLE I.—CORRELATIONS BETWEEN RATE OF WORK AND "COMPREHENSION" MEASURES ON THE IOWA SILENT READING TEST
(N = 110)

Test number	Total time with number attempted in standard time	Total time with score in standard time	Score in standard time with number attempted in standard time
Test 1: paragraph meaning	−.85 ± .02	−.77 ± .03	+.92 ± .01
Test 2: word meaning	−.93 ± .01	−.85 ± .02	+.93 ± .01
Test 3: paragraph organization	−.70 ± .03	−.48 ± .05	+.77 ± .03
Test 4: sentence meaning	−.80 ± .02	−.48 ± .05	+.68 ± .03
Test 5: location of information	−.89 ± .01	−.63 ± .04	+.78 ± .03
Composite total first five tests	−.94 ± .01	−.80 ± .02	+.88 ± .01

measures the working time consumed. Even though some items are done wrong, such is the case in the practical testing situation. The time taken is used for answering items. Some answers are correct; some wrong. We shall take, therefore, the total time used in answering all items of a test as our measure of rate of work.

How adequate a rate of work measure is the number of items attempted in the standard time limit? This question is answered in column two of the table where the coefficients of correlation between number attempted in standard time and total time, our criterion of rate of work, are given. For the various parts of the test, the coefficients, not corrected for attenuation, range from −.70 to −.93. For the test as a whole the correlation was −.94. It is quite evident from these intercorrelations that the number of test items attempted in a set time is a remarkably good measure of rate of work, with the possible exception of Part 3 of the reading test.

The correlations of total time, as a measure of rate of work, and score in standard time as a measure of comprehension are given in column three. The coefficients are of only moderate size for Tests 3, 4, and 5, ranging from −.48 to −.63. For Tests 1 and 2 as well as for the composite score, however, the correlations are fairly high: −.77 to −.85.

In the last column of the table are given the correlations between rate of work and "comprehension" when number attempted in stand-

ard time is used as the measure of rate of work. Again the coefficients are highest for Tests 1 and 2, and the composite score, ranging from +.88 to +.93. Coefficients for the other three tests are moderately high (+.68 to +.78).

It is clear from these results that, for the conditions of this experiment, there is an intimate relationship between rate of work in reading and achievement or comprehension for Tests 1 and 2 and the total comprehension score of the Iowa Silent Reading Test.

Although standardized to measure the reading performance of University freshmen, this test proved fairly easy for University sophomores. Would the correlations have been different if a more difficult test had been used? It is conceivable that the degree to which rate of work is correlated to level of accomplishment in reading is determined by the difficulty of the reading material. To check this suggestion, one should employ our technique for exploring the relationship between rate of work and ability score (speed *vs.* comprehension) in a series of reading tests ranging from very easy to very difficult. Such a program of experiments is now under way in our laboratory.

SUMMARY AND CONCLUSIONS

1. An analysis of the reading situation indicates that the most adequate way to measure the relationship between speed and comprehension in reading is to measure speed and comprehension on the same or strictly comparable material, and then correlate these two scores.

2. Rate of work and ability (comprehension) scores were obtained for the first five tests of the Iowa Silent Reading Test.

3. Number of items attempted in standard time was found to be as good a measure of rate of work as total time taken for the test.

4. Rate of work scores yielded high correlations with comprehension scores which were the time limit scores, especially with tests one, two, and total comprehension (composite of first five tests).

5. The data justify the conclusion that, when an adequate method of measurement is employed, there is an intimate relation between rate of reading and comprehension scores for the type of material here considered. This is contrary to most of the earlier findings.

THE RESIDUAL INDEX*

C. W. YOUNG

Colgate University

INTRODUCTION

The problem is continually arising where a trait Y has been found to correlate with a trait X which is assumed to be a measure of a factor or a series of factors contributing to Y, and it is desired to determine from X and Y a third trait, W, which measures those factors contributing to Y that are not measured by X. The optimal W will not correlate with X and at the same time will correlate with Y to the extent of the coefficient of alienation.

In this paper a formula for calculating such a W; namely, the residual index formula, will be presented, a formula for its reliability derived, an empirical verification of its conformity with the requirements for W as outlined above will be reported, and its relation to the general theory of partial and semi-partial correlation will be pointed out.

THE RESIDUAL INDEX FORMULA

The most general solution of the problem of securing the part of Y that is not related to X is to calculate the Y-residual (or rho) from a line of best fit of Y on X. By definition $(Y - \rho)$ is that part of Y which is functionally dependent upon X, while ρ is the part that is functionally independent.

For the special case in which the relation between X and Y is linear

$$\rho = Y - M_y - b_{yx}(X - M_x) \tag{1}$$

where ρ is the *residual score* of Y on X and the remaining symbols have the usual meanings. In terms of deviations from the means, this formula is

$$\rho = y - b_{yx}x \tag{2}$$

For most work in the field of mental testing, it will not be necessary to calculate the absolute residual scores, but any index which gives a series of values with the same relative standings as ρ will be adequate.

---

* The writer has been greatly assisted in the preparation of this paper by the suggestions and criticisms of Dr. Jack W. Dunlap of Fordham and Professor C. W. Muuıshower of Colgate University.

On the basis of formula (1), such a *residual index* of Y on X may be written

$$RI_{yx} = Y - b_{yx}X \tag{3}$$

since $b_{yx}M_x - M_y$ is constant.

It is worth noting that, since rho is considered error when Y is estimated by X, the standard deviation of the rhos for a series of data is the standard error of estimate of Y on X.

$$\sigma_\rho = \sigma_y\sqrt{1 - r^2} = \sigma_{(est.y)} \tag{4}$$

THE RELIABILITY OF THE RESIDUAL SCORE AND RESIDUAL INDEX

It would be a waste of time to calculate the residual indices for a group of data if the reliability of the indices was so low as to render them worthless. It is, therefore, desirable to predict the reliability of the residual index or residual score from the reliabilities of X and Y and the coefficient of correlation between them. On the basis of equation (2), we may derive the reliability of both the residual score and the residual index as follows:* Let the first measure of rho be

$$\rho_1 = y_1 - b_{yx}x_1 = (y_\infty + e_{y_1}) - b_{yx}(x_\infty - e_{x_1})$$

in which y_∞ and x_∞ are true scores, and e_{x_1} and e_{y_1} are errors of measurement. With a second measure of X and Y we have

$$\rho_2 = y_2 - b_{yx}x_2 = (y_\infty + e_{y_2}) - b_{yx}(x_\infty + e_{x_1})$$

Then

$$\frac{\Sigma_{\rho_1\rho_2}}{N} = \frac{\Sigma(y_\infty + e_{y_1} - b_{yx}x_\infty - b_{yx}e_{x_1})(y_\infty + e_{y_2} - b_{yx}x_\infty - b_{yx}e_{x_1})}{N}$$

On the assumption that the errors are random, this reduces to

$$\frac{\Sigma_{\rho_1\rho_2}}{N} = \frac{\Sigma y^2{}_\infty}{N} + \frac{b^2{}_{yx}\Sigma x^2{}_\infty}{N} - \frac{2b_{yx}\Sigma x_\infty y_\infty}{N}$$

And

$$r_{\rho_1\rho_2}\sigma_{\rho_1}\sigma_{\rho_2} = \sigma^2{}_{y\infty} + b^2{}_{yx}\sigma^2{}_{x\infty} - 2b_{yx}r_{x\infty y\infty}\sigma_{x\infty}\sigma_{y\infty}$$

Assuming that $\sigma_{\rho_1} = \sigma_{\rho_2}$

$$r_{\rho_2\rho_1}\sigma^2{}_\rho = r_{y_1y_2}\sigma^2{}_y + b^2{}_{yx}r_{x_1x_2}\sigma^2{}_x - 2r_{xy}b_{yx}\sigma_x\sigma_y$$

---

* This derivation follows closely Shuttleworth's derivation of the reliability of increments.[4]

By equation (4)

$$r_{\rho_1\rho_2} = \frac{r_{y_1y_2}\sigma^2{}_y + r^2{}_{yx}r_{x_1x_2}\sigma^2{}_y - 2r^2{}_{xy}\sigma^2{}_y}{\sigma^2{}_y(1 - r^2{}_{x_1x_2})}$$

Which reduces to

$$r_{\rho_1\rho_2} = \frac{r_{y_1y_2} - r^2{}_{xy}(2 - r_{x_1x_2})}{1 - r^2{}_{xy}} \qquad (5)$$

Since all r's are positive and fractional, as r_{xy} increases, the reliability of the residual score or residual index decreases. Whenever r_{xy} approaches in magnitude the reliability coefficients of X and Y, it will be well to check by formula (5) to determine whether the residual indices of the scores are worth calculating.

THE USE OF THE RESIDUAL INDEX IN CONSTRUCTING PREDICTIVE TESTS

In the construction of test batteries for predictive purposes, it usually happens that the criterion is related to several factors which cannot all be measured by a single test or type of test. Thus, it is generally recognized that intelligence tests measure only in part the factors which contribute to school or college. "Attitude" describes another group of factors which make significant contributions to scholastic success. But if we validate a test of attitudes against grade point average, we are likely to find that we have simply produced another test of intelligence (since apparently it is easier to find attitudes which correlate with intelligence than to find attitudes which correlate with scholastic success, intelligence eliminated). Thus, Shuttleworth[3] found that an attitudes test which he validated on the basis of grade points at the University of Iowa, correlated approximately .45 with grade point average, but .61 with a composite of entrance examinations. In effect, he had succeeded merely in remeasuring what was already being measured by the tests at hand. A similar instance is found in Chamber's study[1] wherein he validated items on the Pressey X-0 test on the basis of grades. Scoring the X-0 on this basis, he secured the following set of correlations:

Grades and X-0 scores	.46
Grades and intelligence	.53
X-0 and intelligence	.51

Here, again, the correlation with intelligence is higher than the correlation with the criterion.

In order to escape this difficulty, Symonds[5] subtracted the *entire* standard score for intelligence from the standard score for high-school achievement. By doing so he secured an index of "studiousness" which correlated negatively with intelligence, the coefficient being −.56. Using this studiousness index as a criterion, a questionnaire was validated which correlated −.337 with intelligence and .121 with grades. The low correlation with grades doubtless resulted from the fact that the criterion correlated negatively with intelligence.

We may conclude that it is desirable to eliminate from one's criterion of validity that part, but only that part, which is measured by tests already constructed and to use the obtained residual indices as a new criterion; whereupon, one may hope to secure new tests which will combine with the ones at hand to yield higher multiple correlations with the total criterion than could be secured if the latter were employed for the validation of items in the new tests. The method has the additional advantage of supplying a system of uncorrelated measures, all of which correlate with the criterion, thus permitting an analysis of an individual to determine the special factors in which he is lacking. The residual index thus has a wide range of usefulness for test construction throughout the fields of industrial and educational psychology.

AN EMPIRICAL VERIFICATION OF THE RESIDUAL INDEX

Recently the writer, in collaboration with G. H. Estabrooks, made just such a study of the attitudinal factors contributing to scholastic success as has been suggested above. The residual of Y on X was called "studiousness," and a "studiousness index" was calculated by the following formula

$$SI = z_G - z_I r_{GI} \tag{6}$$

where z_G is the standard score for grades, z_I, the standard score for intelligence, and r_{GI}, the coefficient of correlation between grades and intelligence. At the time the writer had not yet hit upon the more convenient residual index formula. The SI, however, is simply the Y-residual in terms of standard units, since, if, from equation (2)

$$\frac{\rho}{\sigma_y} = \frac{y - r_{xy}\frac{\sigma_y}{\sigma_x}x}{\sigma_y}$$

then

$$\frac{\rho}{\sigma_y} = z_y - z_x r_{xy} = SI$$

The results of our investigation of studiousness will be published shortly. Suffice it to say that it was possible, by using the *SI* as a criterion, to develop keys for scoring attitudes questionnaires which correlated approximately zero with intelligence and showed appreciable correlations with grades. In the course of this study the fact that the residual index (or rather its equivalent, the *SI*) is uncorrelated with X and correlated with Y to the extent of the coefficient of alienation of Y on X was empirically verified, since, for a total of five hundred eighty-eight cases, *SI* was found to correlate $-.01$ with intelligence and .94 with grade point average. The coefficient of alienation, calculated from the correlation between grades and intelligence, was .95.

THE RELATION OF THE RESIDUAL TO PARTIAL AND SEMI-PARTIAL CORRELATION

The concept of the residual is basic to the theory of partial and semi-partial correlation. An ordinary partial correlation coefficient, $r_{12.3}$, is essentially the correlation between the residuals of x_1 and x_2 on x_3. Dunlap and Cureton[2] have shown how various systems of semi-partial correlation may be worked out, simply by determining the relations between various combinations of residual and whole scores. At the bottom of page 664 of their article, an equivalent of equation (2) of the present paper appears as a basic assumption for working out partial and semi-partial coefficients.

It follows that whenever one wishes to do no more with a residual trait than to determine its correlation with other traits, residual or otherwise, it will frequently be more economical to secure the coefficient from some formula for partial or semi-partial correlation. In other cases, however, economy will lie in the direction of calculating individual residuals.

Suppose, for example, one wished to determine the correlation of the residual of x_1 on x_3 with x_2. It would only be necessary to use system *B* of Dunlap and Cureton, to secure $r_{(1.3)2}$. Similarly, one might find the correlation between one residual trait and a second by employing Dunlap and Cureton's system *E* (formula 31, p. 670); but in this case, the calculation of the residual scores and the correlation between them would probably prove more economical than the work of calculating all six correlation coefficients required by this particular system of semi-partial correlation.

The residual index will be of use, therefore, when it is desirable to deal with the residual trait in some other way than simply to find its

correlation with some other trait (*e.g.*, as a criterion in item analysis) or when the method of calculating residuals is more economical than the corresponding method of semi-partial correlation.

BIBLIOGRAPHY

1. Chambers, O. R.: "Character trait tests and the prognosis of college achievement." *J. Abn. and Soc. Psychol.*, Vol. XX, 1925, pp. 303–331.
2. Dunlap, Jack W. and Cureton, Edward E.: "On the analysis of causation." *J. Educ. Psychol.*, Vol. XXI, 1930, pp. 657–680.
3. Shuttleworth, F. K.: "Environmental and character factors involved in scholastic success: 1926–1927 data." *J. Educ. Psychol.*, Vol. XX, 1929, pp. 424–433.
4. ———: "The reliability of increments." *J. Educ. Psychol.*, Vol. XXVI, 1935, pp. 312–313.
5. Symonds, P. M.: "A studiousness questionnaire." *J. Educ. Psychol.*, Vol. XIX, 1928, pp. 152–167.

RACIAL COMPARISONS IN PERFORMANCE ON THE AMERICAN COUNCIL PSYCHOLOGICAL EXAMINATION

T. M. LIVESAY

University of Hawaii

This study is concerned with the differences in abilities of four racial groups as shown by performance on a standard college aptitude test. The problem resolves itself into a comparative analysis of the various test scores with appropriate statistical procedure to determine whether revealed differences are significant.

The four racial groups, Caucasian, Chinese, Japanese, and Part-Hawaiian, were used as they constitute the bulk of the undergraduate student body at the University of Hawaii where the study was made. The proportions for the year 1934–1935 were as follows: Japanese thirty-six per cent, Caucasian twenty-four per cent, Chinese twenty-two per cent, Part-Hawaiian ten per cent, and all others eight per cent.

The basic data are the subtest and total scores on the American Council Psychological Examination.[1] This examination requires sixty-two minutes of working time, is easily scored in a few minutes, has an average validity coefficient of .49, and reliability coefficients as follows: Completion .81, Artificial Language .98, Analogies .857, Arithmetic .817, Opposites .87, and Total Score .95.[2]

In order to have numbers adequate for statistical purposes all scores on the 1931, 1932, and 1933 editions, given to freshmen applicants in 1932, 1933, and 1934, were combined and means computed for each racial group on this basis. The assumption is made, of course, that this procedure is equally fair to all groups, but there seems no logical objection as the small differences in scores between the different editions would operate equally on each group and the numbers were in about the same proportion for each year.

In Table I are to be found the basic data—numbers, means, sigmas of the distributions, and sigmas of the means—for the computations as to true differences in Table III. The numbers are not repeated for the various tests as they are naturally the same for each section of the examination.

Table II gives the rank of each race in the different tests and Total Score. The Caucasian group is first in Completion, Analogies,

[1] Prepared by L. L. and T. H. Thurstone, and published by the American Council on Education, 744 Jackson Place, Washington, D. C.

[2] Manual of Instructions, 1933 Edition, pp. 2–4.

Opposites, and Total Score, second in Arithmetic, and third in Artificial Language; the Chinese first in Artificial Language and Arithmetic, second in Analogies, Opposites, and Total Score, and third in Completion; the Japanese second in Artificial Language, third in Analogies, Arithmetic, Opposites, and Total Score, and fourth in Completion;

TABLE I.—NUMBERS, MEANS, SIGMAS OF THE DISTRIBUTIONS, AND SIGMAS OF THE MEANS OF THE FOUR RACIAL GROUPS FOR THE FIVE SUB-TESTS AND THE TOTAL SCORES ON THE AMERICAN COUNCIL PSYCHOLOGICAL EXAMINATION

Test	Measure, N[1]	Caucasian, 265	Chinese, 178	Japanese, 292	Part-Hawaiian, 97
Completion	mean	32.85	25.34	22.82	25.71
	$\sigma_{distribution}$	12.645	11.500	10.450	10.260
	σ_{mean}	.777	.862	.612	1.042
Artificial language	mean	35.83	38.46	36.49	30.87
	$\sigma_{distribution}$	17.385	17.645	16.075	13.480
	σ_{mean}	1.068	1.323	.941	1.369
Analogies	mean	29.57	24.72	24.26	21.12
	$\sigma_{distribution}$	13.880	14.100	14.015	12.845
	σ_{mean}	.853	1.057	.820	1.304
Arithmetic	mean	24.85	25.40	24.09	17.93
	$\sigma_{distribution}$	13.205	15.610	12.750	11.175
	σ_{mean}	.811	1.170	.746	1.135
Opposites	mean	39.70	32.51	31.50	30.20
	$\sigma_{distribution}$	15.960	15.195	14.925	15.895
	σ_{mean}	.980	1.139	.873	1.614
Total score	mean	162.10	146.97	145.03	126.07
	$\sigma_{distribution}$	53.175	52.290	43.440	38.460
	σ_{mean}	3.205	3.919	2.542	3.905

[1] The numbers are the same for each test.

TABLE II.—COMPARISONS OF THE FOUR RACIAL GROUPS BY RANK IN TERMS OF THE MEANS

Race	Completion, rank	Artificial language, rank	Analogies, rank	Arithmetic, rank	Opposites, rank	Total score, rank
Caucasian	1	3	1	2	1	1
Chinese	3	1	2	1	2	2
Japanese	4	2	3	3	3	3
Part-Hawaiian	2	4	4	4	4	4

TABLE III.—COMPARISONS OF THE FOUR RACIAL GROUPS IN TERMS OF THE CHANCES OF A TRUE DIFFERENCE

Race combinations	Tests	Difference between means, D	$\sigma_{(diff.)}$	Critical ratio $\frac{D}{\sigma_{(diff.)}}$	Chances in one hundred that the true difference is greater than zero[1]	Difference in favor of
Caucasian Chinese	Completion	7.51	1.160	6.47	100	Caucasian
	Artificial language	2.63	1.700	1.55	93	Chinese
	Analogies	4.85	1.357	3.57	100	Caucasian
	Arithmetic	.55	1.423	.39	65	Chinese
	Opposites	7.19	1.500	4.79	100	Caucasian
	Total score	15.13	5.060	2.99	100	Caucasian
Caucasian Japanese	Completion	10.03	.989	10.15	100	Caucasian
	Artificial language	.66	1.423	.46	67	Japanese
	Analogies	5.31	1.183	4.49	100	Caucasian
	Arithmetic	.76	1.102	.69	76	Caucasian
	Opposites	8.20	1.312	6.25	100	Caucasian
	Total score	17.07	4.091	4.17	100	Caucasian
Caucasian Part-Hawaiian	Completion	7.14	1.299	5.49	100	Caucasian
	Artificial language	4.96	1.736	2.85	100	Caucasian
	Analogies	8.45	1.558	5.42	100	Caucasian
	Arithmetic	6.92	1.395	4.96	100	Caucasian
	Opposites	9.50	1.888	5.03	100	Caucasian
	Total score	36.03	5.051	7.13	100	Caucasian
Chinese Japanese	Completion	2.52	1.056	2.39	99	Chinese
	Artificial language	1.97	1.623	1.21	88	Chinese
	Analogies	.46	1.337	.34	64	Chinese
	Arithmetic	1.31	1.387	.94	83	Chinese
	Opposites	1.01	1.435	.70	76	Chinese
	Total score	1.94	4.671	.42	65	Chinese
Chinese Part-Hawaiian	Completion	.37	1.352	.27	60	Part-Hawaiian
	Artificial language	7.59	1.903	3.99	100	Chinese
	Analogies	3.60	1.678	2.15	98	Chinese
	Arithmetic	7.47	1.631	4.58	100	Chinese
	Opposites	2.31	1.975	1.17	87	Chinese
	Total score	20.90	5.532	3.78	100	Chinese
Japanese Part-Hawaiian	Completion	2.89	1.208	2.39	99	Part-Hawaiian
	Artificial language	5.62	1.661	3.38	100	Japanese
	Analogies	3.14	1.540	2.04	98	Japanese
	Arithmetic	6.16	1.358	4.54	100	Japanese
	Opposites	1.30	1.835	.71	76	Japanese
	Total score	18.96	4.659	4.07	100	Japanese

[1] Garrett, Henry E.: *Statistics in Psychology and Education.* Table XIV, p. 134.

and the Part-Hawaiian second in Completion, and fourth in all the other parts. Combinations of ranks as well as gradation of total scores place the groups in the following order: Caucasian, Chinese, Japanese, and Part-Hawaiian.

While there is considerable difference between the highest and lowest means, relatively small differences are to be found between some of the others. The question immediately arises as to whether these differences are significant or merely due to chance, and the answer is to be found in the critical ratios in Table III.

In Table III are presented the various race combinations, the differences between the means of the sub-tests and total scores, the sigmas of the differences, the critical ratios for the differences, the chances in one hundred that the true differences are greater than zero, and the racial group in each combination which has the higher mean.

It should be kept in mind that these are group differences, and that any significance which they may have would apply only to these groups in the Territory of Hawaii and continental United States. The Chinese in the United States (Hawaii included) came almost entirely from Canton district and the Japanese mainly from Hiroshima and Yamaguchi Prefectures, and cannot therefore be considered as necessarily representative of the Chinese or Japanese as a whole.

SUMMARY AND CONCLUSIONS

1. This study presents data for two hundred sixty-five Caucasians, one hundred seventy-eight Chinese, two hundred ninety-two Japanese, and ninety-seven Part-Hawaiians on racial differences in performance on the five sub-tests and total scores of the American Council Psychological Examination for the 1931, 1932, and 1933 Editions used as an admissions criterion at the University of Hawaii.

2. The Caucasian group is superior to the other three groups in Completion, Analogies, Opposites, and Total Score; superior to the Japanese and Part-Hawaiian in Arithmetic; and superior only to the Part-Hawaiian group in Artificial Language.

3. The Chinese group is superior to the other three groups in Artificial Language and Arithmetic; superior to the Japanese and Part-Hawaiian in Analogies, Opposites, and Total Score; and superior only to the Japanese in Completion.

4. The Japanese group is superior to the Caucasian and Part-Hawaiian in Artificial Language, and superior only to the Part-Hawaiian in Analogies, Arithmetic, Opposites, and Total Score.

5. The Part-Hawaiian group is superior to the Chinese and Japanese in Completion and inferior to all the others in everything else.

6. The critical ratios indicate complete reliability of the differences in twenty of the thirty-six comparisons.

BOOK REVIEWS

CHARLES E. SKINNER, Editor. *Readings in Psychology.* New York: Farrar and Rinehart, Inc., 1935, pp. XI + 853.

There are a large number of instructors who believe that students in courses in Elementary Psychology should be given outside reading. With this general principle there can be little dispute. Sometimes such outside reading is to other textbooks and sometimes to either journal or monographic literature. A number of people have made collections of readings extracted from the literature with the apparent purpose of relieving the instructor of making up his own reading list. Most of the books of readings have presented rather lengthy extracts, usually sufficient to give the student some contact with a more technical sort of literature than he finds in his textbook. Books of this sort undoubtedly have some usefulness.

Dr. Skinner has edited a new type of book of readings. He has had twenty-five collaborators select materials in some twenty-six different fields of general psychology. Each collaborator has had the responsibility for what should be included in his own chapter. The book is arranged in more or less the traditional fashion in the sequence of topics considered. Instead of preparing lengthy extracts from which the student could get something of the general style of the writers whom they would otherwise never meet, this book has a large proportion of very short bits on very specific points. To the present reviewer, this seems to be an undesirable method.

To quote a rather typical example. On page 159, the following sentence is quoted from Adler with a heading of A Classification of Personality Problems: "Individual psychology has found no problems in life which can not be grouped under three main problems—occupational, social, sexual." Certainly, an extract of this sort gives a very incomplete and erroneous idea of the classification of personality problems to the elementary student. If this were a unique example, there could be no argument, but it is only one of many. Another thing the value of which might be questioned is the very frequent use of extracts from current elementary and, sometimes, more advanced texts which are readily available in practically all university or college libraries. It would seem far better to have the student read the original book rather than an extract or condensation prepared by a second person.

It must be pointed out that the editor and his collaborators have done a good job in making the many small abstracts fit into a reason-

ably adequate sequence. In fact, if the many readings were omitted, one could hardly distinguish the book from another elementary text. As a further attempt to make the book useable as a text, each chapter has been supplied with further reading references and lists of questions and exercises. It is the reviewer's opinion that this might well serve as a basic text in courses in elementary psychology, but is only fairly useful for outside readings. C. M. LOUTTIT.

Indiana University.

JOHN MORRIS DORSEY. *The Foundations of Human Nature.* New York: Longmans, Green and Co., 1935, pp. 488.

In the opening chapters of this book, which carries the subtitle "The Study of the Person," Dr. Dorsey lays a foundation for the understanding of personality. He writes about the springs of human action, mental hygiene insights, perspective, man's biological background, and so on. Following these early chapters the bulk of the book is a delineation of the way one can bring himself to grow toward wholesome living. His treatment of fear, guilt, complexes, religion, happiness, work, play, and self-respect are forcible. He discusses the neglect of the emotional life of the child in the usual educative process. There is a splendid chapter on "Orientation within Personality." Dr. Dorsey reaches out into literature for quotations, and he uses them in an able fashion. The chapter title "Personality in Relation to the World" is indicative of the breath of the whole book, and the currents that are brought together (not always so skillfully) tell us about normal conduct, behavior, where it comes from, and how to attain it.

Foundations of Human Nature can appropriately be called "an accumulation of wisdom, psychiatrically tested and classified, from the humane viewpoint." The book is rather wordy, and Dr. Dorsey has labored over some of his materials. He could have cut through some of the discussion he presented to get at a better organization of certain parts of the book. The author, a psychiatrist, is an officer of a clinic and a consultant for University of Michigan students. He has written this book specifically for educators, but it will be of interest to personnel officers, social workers, and others who are trying to know something of human nature. He has produced a book of considerable meat. It is not easy reading for the person who had had only a bowing acquaintance with psychology or the field of mental hygiene. A carefully prepared summary of pertinent points or terse conclusions terminates each chapter. FORREST H. KIRKPATRICK.

Bethany College, West Virginia.

PUBLICATIONS RECEIVED

AUGUST AICHHORN. *Wayward Youth.* New York: The Viking Press, 1935, pp. 236.

RAYMOND R. AMMARELL. *Workbook and Study Outline for Problems of American Democracy.* Philadelphia: McKinley Pub. Co., 1935, pp. 88. (paper.)

LAWRENCE AUGUSTUS AVERILL. *Adolescence: A Study in the Teen Years.* Boston: Houghton Mifflin Co., 1936, pp. 496.

G. W. BAEHNE, *Editor. Practical Applications of the Punched Card Method in Colleges and Universities.* New York: Columbia University Press, 1935, pp. xxii + 442.

HARRY J. BAKER and VIRGINIA TRAPHAGEN. *The Diagnosis and Treatment of Behavior-problem Children.* New York: The Macmillan Co., 1935, pp. 393.

JOHN EDWARD BENTLEY. *Problem Children.* New York: W. W. Norton & Co., Inc., 1936, pp. 437.

SMILEY BLANTON and MARGARET GRAY BLANTON. *For Stutterers.* New York: D. Appleton-Century Co., 1936, pp. 191.

WALTER BLUMENFELD. *Jugend als Konflikt-situation.* Berlin, Germany: Philo Verlag, 1936. (paper.)

CHARLOTTE BUEHLER and HILDEGARD HETZER. *Testing Children's Development from Birth to School Age.* Translated from the German by Henry Beaumont-New York: Farrar and Rinehart, Inc., 1935, pp. 191.

G. T. BUSWELL. *How People Look at Pictures.* Chicago: University of Chicago Press, 1935, pp. xv + 198.

NELLIE M. CAMPBELL. *The Elementary School Teacher's Treatment of Classroom Behavior Problems.* New York: Bureau of Publications, Teachers College, Columbia University, 1935.

F. STUART CHAPIN. *Contemporary American Institutions: A Sociological Analysis.* New York: Harper & Brothers, 1936, pp. 423.

CLARA FRANCES CHASSELL. *The Relation between Morality and Intellect.* New York: Bureau of Publications, Teachers College, Columbia University, 1936, pp. 556.

HAROLD COE COFFMAN. *American Foundations: A Study of Their Rôle in the Child Welfare Movement.* New York: Association Press, 1936, pp. 213.

LUELLA COLE. *Psychology of Adolescence.* New York: Farrar & Rinehart, Inc., 1936.

MARY COLLINS and JAMES DREVER. *Psychology and Practical Life.* London: University of London Press, 1936.

DENNIS H. COOKE. *Minimum Essentials of Statistics.* New York: The Macmillan Co., 1936, pp. 271.

JOHN FRANCIS CRAMER. *Australian Schools through American Eyes.* Melbourne: Melboune University Press, 1936, pp. 59. (paper.)

BESS V. CUNNINGHAM. *Family Behavior.* Philadelphia: W. B. Saunders Co., 1936, pp. 471.

JOHN EISELE DAVIS and WILLIAM RUSH DUNTON, JR. *Principles and Practice of Recreational Therapy for the Mentally Ill.* New York: A. S. Barnes and Co., 1936, pp. 206.

FLOYD C. DOCKERAY. *General Psychology.* Revised Edition. New York: Prentice-Hall, Inc., 1935.

DOM PROFACE. *College Men: Their Making and Unmaking.* New York: P. J. Kenedy and Sons, 1935, pp. 314.

ANITA SHEMWELL DOWELL. *The Physical Disability of Teachers in the White Elementary Schools of Baltimore, Maryland.* Baltimore: Johns Hopkins Press, 1936, pp. 98. (paper.)

Education for Citizenship in Secondary Schools. Issued under the auspices of The Association for Education in Citizenship. New York: Oxford University Press, 1936, pp. 263.

NORMAN FENTON. *The Delinquent Boy and the Correctional School.* Claremont, Calif.: Claremont College Guidance Center, 1935, pp. 182. (paper.)

PAUL F. FINNER. *An Introduction to Experimental Psychology.* New York: Prentice-Hall, Inc., 1935, pp. 268.

LILLIAN MOHR FOX and L. THOMAS HOPKINS. *Creative School Music.* Newark, N. J.: Silver Burdett Co., 1936, pp. 326.

DOUGLAS FRYER and EDWIN R. HENRY. *Outline of General Psychology.* New York: Barnes & Noble, Inc., 1936, pp. xliii + 229. (paper.)

HAROLD V. GASKILL. *Personality.* New York: Prentice-Hall, Inc., 1936, pp. 52. (paper.)

ARTHUR I. GATES. *Generalization and Transfer in Spelling.* New York: Bureau of Publications, Teachers College, Columbia University, 1935, pp. 80.

SISTER M. XAVIER HIGGINS. *Reducing the Variability of Supervisors' Judgment.* Baltimore: The Johns Hopkins Press, 1936. (paper.)

GLENN D. HIGGINSON. *Psychology.* New York: The Macmillan Co., 1936, pp. 646.

CHRISTINE P. INGRAM. *Education of the Slow-learning Child.* Yonkers, N. Y.: World Book Co., 1935, pp. 419.

NATHAN ISRAELI. *Outlook upon the Future of British Unemployed, Mental Patients, and Others.* Lancaster, Penna., The Science Press Printing Co., 1935, pp. 30. (paper.)

ARTHUR T. JERSILD and SYLVIA F. BIENSTOCK. *Development of Rhythm in Young Children.* New York: Bureau of Publications, Teachers College, Columbia University, 1935. (paper.)

I. L. KANDEL. *Educational Yearbook of the International Institute of Teachers College, Columbia University*—1935. New York: Bureau of Publications, Teachers College, Columbia University, 1935, pp. 631.

ROSCOE GEORGE LINDER. *An Evaluation of Courses in Education of a State Teachers College by Teachers in Service.* New York: Bureau of Publications, Teachers College, Columbia University, 1935.

H. T. MANUEL. *Spanish and English Editions of the Stanford-Binet in Relation to the Abilities of Mexican Children.* Austin: The University of Texas, 1935, pp. 63. (paper.)

HARRY C. MCKOWN. *Character Education.* New York: McGraw-Hill Book Co., Inc., 1935, pp. 472.

A. GORDON MELVIN. *The Activity Program.* New York: Reynal & Hotchcock, 1936.

ELMER D. MITCHELL, *Editor. Sports for Recreation and How to Play Them.* New York: A. S. Barnes and Co., 1936, pp. 467.

JAY B. NASH. *Professional Preparation.* New York: A. S. Barnes and Co., Inc., 1935, pp. 423.

MARTIN H. NEUMEYER and ESTHER S. NEUMEYER. *Leisure and Recreation.* New York: A. S. Barnes and Co., 1936, pp. 405.

LOUIS V. NEWKIRK and HARRY A. GREENE. *Tests and Measurements in Industrial Education.* New York: John Wiley & Sons, Inc., 1935, pp. 253.

B. PETERMANN and KARL HAGGE. *Gewachfene Raumlehre.* Freiburg, Germany: Herder and Co., 1935, pp. 196. (paper.)

HENRY D. RINSLAND. *Analysis of Completion Sentences and Arithmetical Problems as Items for Intelligence Tests.* New York: Bureau of Publications, Teachers College, Columbia University, 1935.

JOSEPH ROEMER, CHARLES FORREST ALLEN, and DOROTHY ATWOOD YARNELL. *Basic Student Activities.* Newark, N. J.: Silver, Burdett and Co., 1935, pp. 367.

CHRISTIAN A. RUCKMICK. *The Psychology of Feeling and Emotion.* New York: McGraw-Hill Book Co., Inc., 1936, pp. 529.

CARL E. SEASHORE. *Psychology of the Vibrato in Voice and Instrument.* Studies in the Psychology of Music, Volume III. Iowa City: The University Press, 1936, pp. 159. (paper.)

GEORGE SETH and DOUGLAS GUTHRIE. *Speech in Childhood: Its Development and Disorders.* New York: Oxford University Press, 1935, pp. 224.

LAURANCE FREDERIC SHAFFER. *The Psychology of Adjustment: An Objective Approach to Mental Hygiene.* Boston: Houghton Mifflin Co., 1936.

CHARLES E. SKINNER, *Editor. Readings in Psychology.* New York: Farrar and Rinehart, Inc., 1935, pp. 853.

JOSEF SOMOGYI. *Begabung im Lichte der Eugenik.* Wien, Austria: Franz Deuticke, 1936, pp. 518. (paper.)

HERBERT SORENSON. *Statistics for Students of Psychology and Education.* New York: McGraw-Hill Book Co., 1936, pp. 373.

WILLIAM STERN. *Allgemeine Psychologie auf Personalistischer Grundlage.* The Hague: Martinus Nijhoff, 1935, pp. 831. (paper.)

JOHN W. STUDEBAKER. *Plain Talk.* Washington, D. C.: National Home Library Foundation, 1936, pp. 166.

Studies in Honor of William Torrey Harris on the Occasion of the Hundredth Return of His Birthday, September 10, 1935. Published with the Coöperation of Scholars in America and Abroad by the International Education Review. Berlin: Weidmannsche Buchhandlung, 1935, pp. 316. (paper.)

Teachers' Guide to Child Development in the Intermediate Grades. Prepared under the Direction of the California State Curriculum Commission. Sacramento: California State Department of Education, 1936, pp. 631.

PERCIVAL M. SYMONDS. *Education and the Psychology of Thinking.* New York: McGraw-Hill Book Co., Inc., 1936, pp. 306.

WILLARD LEE VALENTINE and OTHERS. *Student's Guide for Beginning the Study of Psychology.* New York: Prentice-Hall, Inc., 1935, pp. 267. (paper.)

WAYLAND F. VAUGHAN. *General Psychology.* Garden City, N. Y.: Doubleday, Doran and Co., Inc., 1936.

HELEN M. WALKER and WALTER N. DUROST. *Statistical Tables: Their Structure and Use.* New York: Bureau of Publications, Teachers College, Columbia University, 1936, pp. 76.

J. E. WALLACE WALLIN. *Personality Maladjustments and Mental Hygiene.* New York: McGraw-Hill Book Co., Inc., 1935, pp. 511.

ELIZABETH WATERMAN. *Rhythm Book: A Manual for Teachers of Children.* New York: A. S. Barnes and Co., 1936, pp. 150.

CARL HAVELOCK WEDELL. *Workbook in Psychology.* New York: John Wiley and Sons, Inc., 1936, pp. 129. (paper.)

THEODORE WEISENBURG, ANNE ROE, and KATHARINE E. MCBRIDE. *Adult Intelligence.* New York: The Commonwealth Fund, 1936. (paper.)

FRANK T. WILSON. *Guiding Our Children.* New York: Globe Book Co., 1935, pp. 248.